



Ladislaus von Bortkiewicz Lehrstuhl für Statistik  
Institut für Statistik und Ökonometrie  
Wirtschaftswissenschaftliche Fakultät  
Humboldt-Universität zu Berlin

BACHELORARBEIT

ZUR ERLANGUNG DES AKADEMISCHEN GRADES BACHELOR OF SCIENCE  
IN VOLKSWIRTSCHAFTSLEHRE

Eine statistische Analyse zur Situation von Frauen in Deutschland

A statistical analysis of the women in Germany



Bearbeiterin:	Shujun Huang
Matrikelnummer:	544154
Abgabedatum:	31.08.2015
Prüfer:	Prof. Dr. Wolfgang Härdle
Betreuer:	Dr. Sigbert Klinke

# Inhaltsverzeichnis

<b>1. Kapitel: Einleitung</b>	4
<b>2. Kapitel: Datengrundlage</b>	5
2.1 Beschreibung der Datenbasis	5
2.2 Datenaufbereitung	6
2.3 Deskriptive Statistik	7
<b>3. Kapitel: Angewandte Methodik</b>	10
3.1 Test auf Verteilungsunterschiede	10
3.1.1 Chi-Quadrat-Test	10
3.1.4 Kruskal-Wallis-Test	13
3.2 Two-Step-Clusteranalyse	14
3.2.1 Ablauf der Two-Step-Clusteranalyse	14
3.2.2 Bestimmung der optimalen Clusterzahl: Bayes'sches Informationskriterium	16
3.2.3 Bewertung der Modellgüte: Silhouette	17
3.3 Korrelationsanalyse	18
3.4 Faktorenanalyse	20
3.4.1 Ablauf der Faktorenanalyse	20
3.4.2 Reliabilitätsanalyse	22
<b>4. Kapitel: Auswertung der Ergebnisse</b>	23
4.1 Nichtparametrische Tests	23
4.1.1 Vergleich der Stichproben aus drei Erhebungsjahren	23
4.1.2 Vergleich der Stichproben aus beiden Erhebungsgebieten	26
4.2 Clusteranalyse	28
4.3 Korrelationsanalyse	29
4.4 Faktorenanalyse	32
<b>5. Kapitel: Fazit</b>	35
<b>Literaturverzeichnis</b>	36
<b>Anhang: Tabellen und Abbildungen</b>	38
<b>Erklärung zur Urheberschaft</b>	54

## Tabellenverzeichnis

Tabelle 1: Variablenlist.....	6
Tabelle 2: Bewertung der Ergebnisse von dem KMO-Maß.....	21
Tabelle 3: Vergleich der Familienstand nach Erhebungsjahr .....	23
Tabelle 4: Bildung, Status der Berufstätigkeit und Einkommen nach Erhebungsjahr.....	23
Tabelle 5: Einstellungen nach Erhebungsjahr.....	26
Tabelle 6: Familienstand nach Erhebungsgebiet .....	27
Tabelle 7: Bildung und Status der Berufstätigkeit nach Erhebungsgebiet.....	27
Tabelle 8: Einkommen nach Erhebungsgebiet .....	28
Tabelle 9: Auszug der Korrelationstabelle .....	30
Tabelle 10: Auszug der Korrelationstabelle.....	31
Tabelle 11: KMO-Maß.....	32
Tabelle 12: erklärte Gesamtvarianz.....	33
Tabelle 13: Komponentenmatrix .....	34
Tabelle 14: Cronbach's Alpha .....	34

## Abbildungsverzeichnis

Abbildung 1: Bildungsgrad.....	7
Abbildung 2: Status der Berufstätigkeit.....	8
Abbildung 3: monatliches Nettoeinkommen.....	9
Abbildung 4: Fehlerbalkendiagramm: Bildungsgrad nach Erhebungsjahr .....	24
Abbildung 5: Fehlerbalkendiagramm: Status der Berufstätigkeit nach Erhebungsjahr .....	25
Abbildung 6: Fehlerbalkendiagramm: Einkommen nach Erhebungsjahr .....	25
Abbildung 7: Screeplot .....	33

# 1. Kapitel: Einleitung

Die Erkenntnis, dass „Mann“ und „Frau“ nicht einfach biologische Tatsachen abbilden, sondern Geschlechter in sozialen Prozessen darstellen, ist eine der zentralen Leistungen der Frauenbewegung in den 70er Jahren. Seitdem wird die Geschichte vom „Ende der Normalfamilie“, in der der Vater die Erwerbsarbeit und die Mutter die unentgeltliche Familienarbeit jeweils allein leisteten, meist als Geschichte der Befreiung der Frauen erzählt. Rückblickend lässt sich die Frage stellen, wie sich die Lebenssituation der Frauen in der Vergangenheit verändert hat. Im Zentrum dieser Arbeit stehen folgende Fragestellungen:

- Wie haben sich die sozioökonomischen Bedingungen der Frauen entwickelt?
- Wie haben sich die Einstellungen der Frauen hinsichtlich Familie und Geschlechterrolle verändert?
- Gibt es verschiedene Teilgruppen von Frauen, die sich anhand ihrer sozioökonomischen Bedingungen und Einstellungen zur Familie und Geschlechterrolle unterscheiden?
- Wodurch werden die Einstellungen der Frauen hinsichtlich Familie und Geschlechterrolle beeinflusst?

Die Untersuchung beschränkt sich in dem Zeitraum von 1988 bis 2012 auf die erwachsenen Frauen in Deutschland.

In dieser Arbeit wird zunächst die verwendete Datengrundlage in Kapitel 2 beschrieben. Danach werden die angewandten Methoden im Kapitel 3 vorgestellt. Im Mittelpunkt des vierten Kapitels steht die Auswertung der Daten. Schlussendlich werden die gefundenen Ergebnisse im Kapitel 5 zusammengefasst.

## 2.Kapitel: Datengrundlage

### 2.1 Beschreibung der Datenbasis

Das Datenmaterial der vorliegenden Arbeit stammt aus den Datensätzen der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaft (ALLBUS). Bei der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaft handelt es sich um eine repräsentative Bevölkerungsumfrage für die erwachsene Wohnbevölkerung in Deutschland, die seit 1980 in Westdeutschland sowie seit 1991 in Gesamtdeutschland in einem Zwei-Jahres-Rhythmus durchgeführt wird.

Das Frageprogramm der ALLBUS-Umfrage wird so gestaltet, dass es zum Teil konstant, zum Teil variabel ist. In dem konstanten Themenblock werden Fragen nach demographischen Merkmalen und einigen grundlegenden gesellschaftlich relevanten Merkmalen in jeder Umfrage gestellt. Der variable Teil besteht aus verschiedenen kleinen Themenblöcken, deren Schwerpunkt im Abstand von Zehn Jahren wiederholt wird. Die langjährige Durchführung der Umfrage bietet die Gelegenheit, Einsicht über die Entwicklung der Lebensverhältnisse und die Veränderung ihrer Wertorientierungen zu erhalten.

Zur Beantwortung der Forschungsfrage kommen in erster Linie die ALLBUS-Daten in den Jahrgängen in Frage, deren Themenschwerpunkt auf der Familien- und Geschlechtsrolle liegt. Da das Ziel dieser Arbeit auf eine möglichst langfristige Entwicklungssituation ausgerichtet ist, sollten Datensätze aus den 80er und 90er Jahren auch herangezogen werden. Schlussendlich wurden die ALLBUS-Daten in den Jahren 1988, 1994 und 2012 als Datenbasis der vorliegenden Arbeit ausgewählt.

## 2.2 Datenaufbereitung

Der erste Schritt der Datenaufbereitung beinhaltet eine Auswahl der relevanten Variablen. Betrachtet werden sollten die Variablen, die sich als Indikatoren für demographische Merkmale, sozioökonomische Bedingungen und subjektive Einstellungen zur Geschlechterrolle und Familie eignen. Um einen Zeitvergleich zu ermöglichen, sollten auch die Variablen ausgewählt werden, die in den Datensätzen der oben genannten drei Erhebungsjahre beinhaltet sind. Nach einer vollständigen Recherche und einem Vergleich der Codebücher wurden die in Tabelle 1 aufgezeigten Variablen in der vorliegenden Analyse verwendet. Sie wurden aus dem originalen Datensatz in dem jeweiligen Jahrgang kopiert und in einem neuen Datensatz zusammengeführt. Um zu identifizieren, aus welchem Jahr die Merkmale erhoben wurden, wurde in dem neuen Datensatz eine neue Variable, „Erhebungsjahr“, für jede Beobachtungseinheit generiert. Als nächster Schritt wurde die Codierung der Variablen aus drei Erhebungsjahren vereinheitlicht. Durch die Umsetzung der Datenaufbereitung entsteht ein veränderter Datensatz, der als Datengrundlage der weiteren Analyse dient.

sozioökonomische Indikatoren	Items über Einstellungen zum Thema
GESCHLECHT, BEFRAGTE<R>	BERUFSTAETIGE FRAU: FAMILIENLEBEN LEIDET
ALTER: BEFRAGTE<R>	
ERHEBUNGSGEBIET	FRAUEN ZIEHEN HEIM UND KINDER BERUF VOR
<WOHNGBIET>: WEST - OST	
FAMILIENSTAND, BEFRAGTE<R>	MANN GELD VERDIENEN, FRAU IM HAUSHALT
ALLGEMEINER SCHULABSCHLUSS	
BEFR.: FACHSCHULABSCHLUSS	BEI KINDERWUNSCH SOLLTE MAN HEIRATEN
BEFR.: HOCHSCHULABSCHLUSS	
BEFRAGTE<R> BERUFSTAETIG?	
BEFR.: NETTOEINKOMMEN,	
OFFENE ABFRAGE	

**Tabelle 1: Variablenlist**

## 2.3 Deskriptive Statistik

In Tabelle 2-1 im Anhang wurde die deskriptive Statistik der Variablen angeführt, die als Indikatoren für die objektiven soziodemographischen Merkmale und die sozio-ökonomischen Rahmenbedingungen dienen.

Die größte Veränderung tritt bei dem Bildungsgrad auf. Anhand der Quantilswerte ist zu sehen, dass die Verteilung des Bildungsniveaus nach oben verschoben wurde. In Abbildung 1 wird verdeutlicht, dass der Anteil der Frauen, die höchstens über einen Hauptschulabschluss verfügen, in dem untersuchten Zeitraum stark reduziert ist. Gleichzeitig ist der Anteil anderer Bildungsgruppen erheblich angestiegen. Insbesondere ist zu sehen, dass sich der Anteil der hoch qualifizierten Frauen fast verdoppelt.

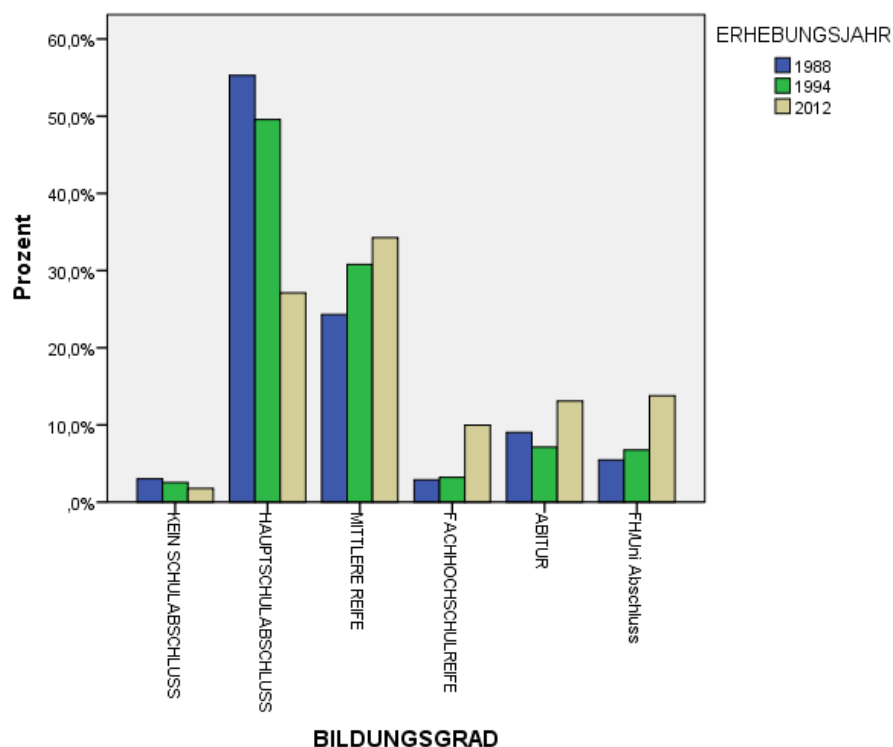
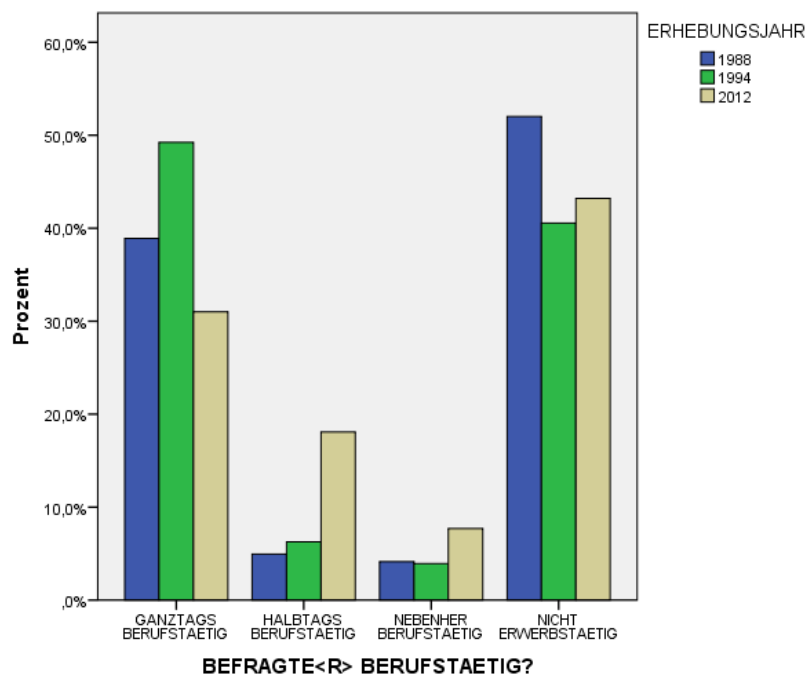


Abbildung 1: Bildungsgrad

Darüber hinaus ist die Veränderung in der Verteilung des Status der Erwerbstätigkeit betrachtenstwert. In dem Balkendiagramm in Abbildung 2 ist zu sehen, dass der Anteil der ganztags erwerbstätigen Frauen und der Anteil der nichterwerbstätigen Frauen im Vergleich zu den früheren Erhebungsjahren im Jahr 2012 gesunken sind. Gleichzeitig sind die Anteile der halbtags erwerbstätigen und nebenher erwerbstätigen Frauen deutlich gestiegen.

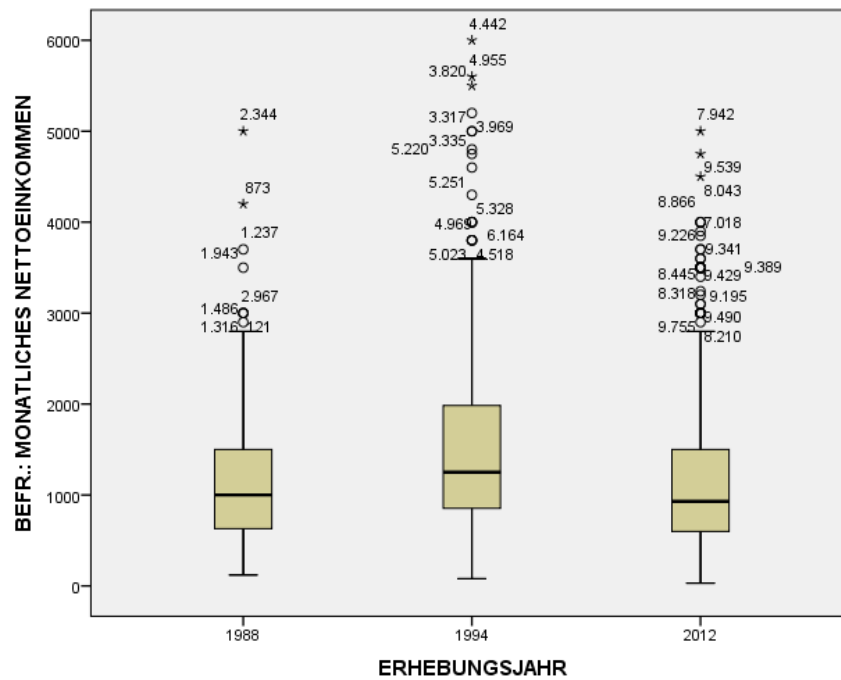


**Abbildung 2: Status der Berufstätigkeit**

Interessanterweise erscheint die Einkommensverteilung im Jahr 2012 ähnlich wie die in den 80er Jahren. Sowohl die Mittel- als auch die Quantilswerte liegen nahe beieinander. Jedoch ist die gesamte Einkommensverteilung in den 90er Jahren höher gelagert als in den anderen beiden Jahren. Dieses Ergebnis wurde mit dem Boxplot in Abbildung 3 betätigt. Außerdem ist aus dem Boxplot zu sehen, dass die Ausreißer im Lauf der Zeit mehr geworden sind.



In Tabelle 2-2 im Anhang ist die deskriptive Statistik der Variablen aufgezeigt, die sich auf die subjektiven Einstellungen der Befragten zur Familien- und Geschlechterrolle beziehen. Es ist deutlich zu sehen, dass die Verteilung aller untersuchten Variablen im Lauf der Zeit nach oben verschoben ist, d. h., im Laufe der Zeit tendieren die Befragten dazu, traditionelle Einstellungen abzulehnen.



**Abbildung 3: monatliches Nettoeinkommen**

## 3.Kapitel: Angewandte Methodik

### 3.1 Test auf Verteilungsunterschiede

#### 3.1.1 Chi-Quadrat-Test

Um zu überprüfen, ob sich die Verteilungen  $k$  unabhängiger Stichproben signifikant voneinander unterscheiden, kann der Chi-Quadrat-Test eingesetzt werden, wenn die Variablen nominal oder ordinal skaliert sind. Hierbei ist die Nullhypothese, dass die  $k$  Stichproben aus einer Grundgesamtheit bzw. Grundgesamtheiten mit identischer Verteilung stammen, durch folgende Prüfgröße geprüft:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(h_{ij} - e_{ij})^2}{e_{ij}} \quad (1)$$

Während  $h_{ij}$  die beobachtete Häufigkeit in der  $i$ -ten Zeile und  $j$ -ten Spalte darstellt, ist  $e_{ij}$  entsprechend die erwartete Häufigkeit unter  $H_0$  in  $i$ -ter Zeile und  $j$ -ter Spalte.<sup>1</sup>

Hinsichtlich der Gültigkeit der Nullhypothese ist die Teststatistik approximativ Chi-verteilt mit einem Freiheitsgrad von  $df = (k - 1)(r - 1)$ , wobei  $k$  für die Anzahl der Spalte und  $r$  die Anzahl der Zeilen steht.

---

<sup>1</sup> Vgl. Siegel, S.(1956): Nonparametric statistics for the behavioral sciences, S.104-105

### 3.1.2 Kolmogorov-Smirnow-Test

Der Kolmogorov-Smirnow-Test stellt ein nichtparametrisches Verfahren dar, welches überprüft, ob das beobachtete Merkmal eine vollständig bestimmte Verteilung verfolgt. Die Grundidee des Kolmogorov-Smirnow-Tests liegt in dem Vergleich der empirischen Verteilung mit der hypothetischen Verteilung. Bei Gültigkeit der Nullhypothese, welche annimmt, dass die beobachtete Verteilungsfunktion eine genau spezifische Gestalt  $F_0$  aufweist, soll die Differenz zwischen der beobachteten und hypothetischen Verteilung für alle  $x$  hinreichend klein sein.

Bei dem Kolmogorov-Smirnow-Test für Zweistichproben wird geprüft, ob die beiden Stichproben aus einer Grundgesamtheit bzw. aus zwei Grundgesamtheiten mit der gleichen Verteilung stammen. Die Voraussetzungen der Stichprobenvariablen sind analog zum Kolmogorov-Smirnow-Test bei einer Einstichprobe. Zusätzlich müssen die beiden Stichproben unabhängig voneinander sein.

Die Testhypothese im Fall Zweistichprobe lautet:

$$H_0: F_X(X) = F_0(X), H_1: F_X(X) \neq F_0(X)$$

Die Teststatistik<sup>2</sup> ist wie folgt definiert:

$$D_n = |F_n(X) - F_0(X)| \quad (2)$$

---

<sup>2</sup> Vgl. Siegel, S.(1956): Nonparametric statistics for the behavioral sciences, S.127-128

Vgl. Büning, H., et al (1994): Nichtparametrische statistische Methoden, S.119-120

### 3.1.3 Mann-Whitney-Test

Bei dem Mann-Whitney-Test handelt es sich um ein verteilungsfreies Testverfahren zum Vergleich zweier unabhängiger Stichproben hinsichtlich ihrer zentralen Tendenz, die die Unabhängigkeit und ein mindestens ordinales Skalenniveau der Stichprobenvariablen voraussetzt. In diesem Testverfahren des Mann-Whitney-Tests werden alle Messwerte der Größe nach geordnet und mit ihrem Rangplatz versehen. Danach werden die Rangsummen der beiden Stichproben miteinander verglichen. Existiert ein Unterschied in den Summen der Rangplätze in den einzelnen Gruppen, deutet es möglicherweise auf einen Unterschied in den Verteilungen der beiden Grundgesamtheiten hin.

Man geht davon aus, dass  $F_1(x + a) = F_2(x)$  gilt. Folgende Hypothesen werden gegeneinander getestet:

$$H_0: a = 0; H_1: a \neq 0$$

Die Teststatistik lautet:

$$V = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 \approx N(m_u; s_u) \quad (3)$$

wobei  $m_u = \frac{n_1 n_2}{2}$ ,  $s_u^2 = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$  und  $R_1$  die Rangsumme der ersten Stichprobe darstellt.<sup>3</sup>

---

<sup>3</sup> Vgl. Büning, H., et al.(1994): Nichtparametrische statistische Methoden, S.200-201

Vgl. Bortz, J., Lienert, et al (1990): Verteilungsfreie Methoden in der Biostatistik,

5.Auflage, S. 146-147

### 3.1.4 Kruskal-Wallis-Test

Der Kruskal-Wallis-Test kann als die Verallgemeinerung des Mann-Whitney-Tests von zwei Stichproben auf  $p$  Stichproben verstanden werden. Hierbei wird geprüft, ob die Stichproben aus derselben Grundgesamtheit bzw. aus Grundgesamtheiten mit derselben Verteilungsform und zentralen Tendenz stammen. Der Kruskal-Wallis-Test setzt voraus, dass die Stichproben sowie die Stichprobenvariablen unabhängig sein müssen. Darüber hinaus muss die Verteilungsfunktion der Stichprobenvariablen stetig sein.

Die Testhypothese des Kruskal-Wallis-Tests lautet:

$$\begin{aligned} H_0: F_1(z) &= F_2(z) = \dots = F_c(z) \\ H_1: F_i(z) &= F(z - \theta_i) \text{ für alle } z \in \mathbb{R} \text{ und mit } \theta_i \neq \theta_j \\ &\text{für mindestens ein Paar } (i, j), 1 \leq i, j \leq c \end{aligned}$$

An dieser Stelle ist zu beachten, dass man bei Ablehnung der Nullhypothese nur davon ausgehen kann, dass mindestens zwei Stichproben hinsichtlich ihrer zentralen Tendenz unterschiedlich sind.

Die Teststatistik des Kruskal-Wallis-Tests ist wie folgt definiert:

$$V = \frac{12}{n(n+1)} \sum_{j=1}^p n_j \left( R_j - \frac{n+1}{2} \right)^2 \approx \chi_{p-1}^2 \quad (4)$$

wobei  $R_j$  die Rangsumme von Gruppe  $j$  ist.<sup>4</sup>

---

<sup>4</sup> Vgl. Siegel, S.(1956): Nonparametric statistics for the behavioral sciences, S.184-185  
Vgl. Büning, H., et al.(1994): Nichtparametrische statistische Methoden, S.200-201

## 3.2 Two-Step-Clusteranalyse

Im Vergleich zu den traditionellen Clusterverfahren ist die Two-Step-Clusteranalyse durch die gleichzeitige Verarbeitung von gemischtskalierten Daten sowie den geringen Rechenaufwand bei großem Datensatz gekennzeichnet. Darüber hinaus ist es im Rahmen einer Two-Step-Clusteranalyse möglich, die optimale Clusterzahl automatisch zu bestimmen.

Die Grundidee der Two-Step-Clusteranalyse besteht darin, durch eine Vorsortierung die rohen Daten in mehrere kleinere Datensätze zu unterteilen und somit das Datenvolumen zu reduzieren. Ausgehend von dieser Überlegung werden zunächst die Daten in verschiedene kleine Untercluster mittels Bildung eines sogenannten CF-Baums eingeteilt, im zweiten Schritt wird dann ein hierarchisches Clusterverfahren auf die Subcluster angewendet.

### 3.2.1 Ablauf der Two-Step-Clusteranalyse

Im ersten Schritt der Two-Step-Clusteranalyse werden zunächst die einzelnen Objekte der Daten eingelesen und mittels der Bildung eines CF-Baums (Cluster Feature Tree, Gruppenmerkmalebaum), die auf der von Zhang et al. (1997) entwickelten BIRCH-Methode basiert, in verschiedene Untercluster sortiert.

Die einzelnen Objekte werden vom Ursprung ausgehend an jedem Knoten zu der Verzweigung weitergeleitet, die am nächsten liegt. Die Messung dieser Nähe bzw. Distanz wird mithilfe des Log-Likelihood-Distanzmaßes berechnet, falls keine kategoriellen Variablen vorhanden sind. In den Endknoten auf der letzten Stufe befinden sich die Untercluster, die durch die bereits eingelesenen Objekte gebildet sind. In solchen Unterclustern sind jedoch keine Objekte anzutreffen, sondern nur die so

genannten Cluster Features. Ein Cluster Features besteht aus statistischen Kennzahlen und charakterisiert das Endcluster. Mittels der Cluster Features kann die Distanz zwischen zwei Clustern berechnet werden, ohne die genauen Objekte kennen zu müssen.

Die Log-Likelihood-Distanz, die auf Wahrscheinlichkeitsverteilungen basiert, hängt von der Abnahme in der Log-Likelihood-Funktion bei dem Zusammenfassen zweier Cluster ab. Die Distanz zwischen Cluster  $j$  und  $s$  ist definiert durch:

$$d(i, j) = \xi_i + \xi_j - \xi_{\langle i, j \rangle} \quad (5)$$

$$\text{wobei } \xi_v = -N_v \left( \sum_{k=1}^{K^A} \frac{1}{2} \log(\hat{\sigma}_k^2 + \hat{\sigma}_{vk}^2) + \sum_{k=1}^{K^B} \hat{E}_{vk} \right)$$

$$\text{mit } \hat{E}_{vk} = -\sum_{l=1}^{L_k} \frac{N_{vkl}}{N_v} \log \frac{N_{vkl}}{N_v}$$

Der Cluster Feature  $CF_i$  des Clusters  $i$  ist definiert durch

$$CF_i = (N_i, \bar{x}_i, s_i^2, N_{iB}) \quad (6)$$

Bei dem Zusammenfassen zweier Cluster  $i$  und  $j$  lässt sich der neue Cluster Feature  $CF_{\langle i, j \rangle}$  wie folgt berechnen:

$$CF_{\langle i, j \rangle} = (N_i + N_j, x_{\langle i, j \rangle}, s_{\langle i, j \rangle}^2, N_{iB} + N_{jB}), \quad (7)$$

$$\text{wobei } x_{\langle i, j \rangle} = \frac{N_i \bar{x}_i + N_j \bar{x}_j}{N_i + N_j} \text{ und } s_{\langle i, j \rangle}^2 = \frac{(N_i - 1)s_i^2 + (N_j - 1)s_j^2 + \bar{x}_i^2 + \bar{x}_j^2}{N_i + N_j - 1} \text{ sind.}$$

Die Größe eines CF-Baums wird durch die vorgegebene Maximalhöhe und maximale Anzahl von Verzweigungen an jedem Baumknoten festgelegt. Der CF-Baum

kann mehrmals gebildet werden, bis die endgültige Struktur ermittelt wurde.<sup>5</sup>

Als zweiter Schritt der Two-Step-Clusteranalyse werden die Untercluster, die durch den CF-Baum gefunden wurden, mittels eines hierarchischen Clusteranalyseverfahrens zusammengefasst. Durch das Zusammenfassen der zwei nächsten Untercluster zu einem neuen Cluster wird die Clusterzahl reduziert.

### 3.2.2 Bestimmung der optimalen Clusterzahl: Bayes'sches Informationskriterium

Am Ende des hierarchischen Clusterverfahrens wird eine Reihe von Clusterlösungen geliefert, deren Clusteranzahl von eins bis n variiert. Zu einer automatischen Bestimmung der optimalen Clusteranzahl wird das Bayes'sche Informationskriterium, ein Maß für die Auswahl und den Vergleich von Modellen, bei dem Two-Step-Clusteranalyse eingesetzt, falls die Anzahl der Cluster nicht vorgegeben ist. Das Bayes'sche Informationskriterium ist definiert als

$$BIC(J) = -2 \sum_{j=1}^J \xi_j + J \left\{ 2P_A + \sum_{pB=1}^{P_B} (M_{pB} - 1) \right\} \ln(N) \quad (8)$$

Zunächst ist das BIC bei jeder Clusterzahl zu berechnen, danach wird das Verhältnis gebildet, bei welchem die Änderungen im BIC durch das Zusammenfassen

---

<sup>5</sup> Vgl. Chiu, T., et al (1999). A Robust and Scalable Clustering Algorithm for Mixed Type Attributes in Large Database Environment.



zweier Cluster im Zähler und die Veränderung im BIC bei den beiden letzten Clustern zu einem einzigen Cluster im Nenner stehen. Das Verhältnis der Änderung im BIC beim Zusammenfassen zweier Cluster ist

$$R_1(J) = \frac{BIC(J) - BIC(J + 1)}{BIC(1) - BIC(2)} \quad (9)$$

Eine relativ große Veränderung in dem BIC-Verhältnis deutet darauf hin, dass an dieser Stelle zwei Cluster zusammengefasst werden, die sich relativ stark voneinander unterscheiden und deshalb nicht zusammengefasst werden sollen.

### 3.2.3 Bewertung der Modellgüte: Silhouette

Die Silhouette, die als Umriss verstanden werden kann, ist ein Maß für den Grad der Konzentration bzw. Separation der Objekte in einem Modell.

Die Distanz zwischen der Beobachtung  $x$  und dem Cluster  $C$  kann wie folgt beschrieben werden:

$$d(x, C) = \frac{1}{n_C} \sum_{c \in C} d(c, x)$$

Angenommen, dass  $x \in A$  und  $B$  die nächsten Cluster zum  $x$  sind

$$d(x, B) = \min_{C \neq A} d(x, C)$$

Die Silhouette ist dann definiert als

$$S(x) = \begin{cases} 0 \\ \frac{d(x, B) - d(x, A)}{\max(d(x, B), d(x, A))} \end{cases} \quad (10)$$

Bei einer guten Clusterlösung soll  $d(x, A)$  möglichst klein sein und  $d(x, B)$  möglichst groß ausfallen soll. Der Wert der Silhouette variiert zwischen -1 und 1.<sup>6</sup>

### 3.3 Korrelationsanalyse

Der Korrelationskoeffizient ist ein Maß für die Stärke und die Richtung des Zusammenhangs zweier Variablen, welches einen Wert von -1 bis +1 annehmen kann. Ein positiver Korrelationskoeffizient deutet darauf hin, dass zwischen den beiden Variablen ein positiver linearer Zusammenhang besteht. Bei einem negativen Korrelationskoeffizienten ist es genau umgekehrt. Je stärker der lineare Zusammenhang zwischen den Variablen ist, desto größer wird der Betrag des Korrelationskoeffizienten.

Der Korrelationskoeffizient ist darauf beschränkt, den linearen Zusammenhang zwischen den Variablen zu identifizieren. Außerdem ist stets zu beachten, dass eine Korrelation keinen Beleg für einen kausalen Zusammenhang zwischen den Variablen ergibt.

Bei mindestens ordinal skalierten Variablen kann der Korrelationskoeffizient von Spearman eingesetzt werden, um die Stärke und die Richtung eines linearen Zusammenhangs zu messen. Der Korrelationskoeffizient von Spearman wird definiert als

---

<sup>6</sup> vgl. SPSS (2001): The SPSS Two Step Cluster Component. White paper – technical report

$$r_s = 1 - \frac{6}{n^3 - n} \sum_{i=1}^n (R(x_i) - R(y_i))^2 \quad (11)$$

wobei  $R(x_i)$  und  $R(y_i)$  für den Rangplatz von  $x_i$  und  $y_i$  stehen.

Bei dem Auftreten von Bindungen wird der Korrelationskoeffizient von Spearman wie folgt berechnet:

$$r_s = \frac{C_x + C_y - \sum_{i=1}^n (R(x_i) - R(y_i))^2}{2\sqrt{C_x C_y}} \quad (12)$$

$$\text{mit } C_x = \frac{n^3 - n}{12} - \frac{1}{12} \sum_k (t_{x,k}^3 - t_{x,k}) \text{ und } C_y = \frac{n^3 - n}{12} - \frac{1}{12} \sum_k (t_{y,k}^3 - t_{y,k})$$

$t_{*,k}$  steht für die Anzahl der Bindungen.

Der Korrelationskoeffizient kann auf einem festgelegten Signifikanzniveau getestet werden. Geprüft werden soll, ob ein wesentlicher Zusammenhang zwischen den beiden Variablen X und Y anzunehmen ist oder ob der geschätzte Wert des Korrelationskoeffizienten als zufallsbedingt angesehen werden muss. Die Null- und Alternativhypothese sind wie folgt formuliert:

$$H_0: r_{yx} = r_0 = 0, \quad H_1: r_{yx} \neq r_0 = 0$$

Ist  $r_{yx} > r_{f,\alpha/2}$ , so liegt auf einem Signifikanzniveau von Alpha und basierend auf einer Stichprobe vom Umfang n ein wesentlicher Zusammenhang zwischen den Variablen Y und X vor. Ist  $r_{yx} < r_{f,\alpha/2}$ , so lässt sich über den untersuchten Zusammenhang aufgrund der Stichprobe nichts aussagen.<sup>7</sup>

---

<sup>7</sup> Vgl. Brosius, F. (2013): SPSS 21, S.523-525

## 3.4 Faktorenanalyse

### 3.4.1 Ablauf der Faktorenanalyse

In der Praxis sind die untersuchten Variablen der Grundgesamtheit oft nicht voneinander unabhängig. Diese Situation lässt sich möglicherweise darauf zurückführen, dass die untersuchten Variablen durch latente Größen beeinflusst sind, die nicht gemessen wurden bzw. nicht direkt gemessen werden können. Solche latenten Variablen, die den wechselseitigen Korrelationen der beobachteten Variablen zugrundeliegen, werden als Faktor bezeichnet. Das Ziel der Faktorenanalyse besteht darin, solche Faktoren zu identifizieren, ihre Existenz zu bestätigen bzw. die Faktorenwerte zu berechnen.

Die Faktorenanalyse geht davon aus, dass wechselseitige Korrelationen zwischen den beobachteten Variablen existieren. Aber die Korrelationsmatrix selbst lässt noch kein eindeutiges Urteil über die Eignung der Daten für die Faktorenanalyse zu. Deshalb ist es empfehlenswert, die Korrelationskoeffizienten vorab auf ihre Eignung zur Faktorenanalyse zu überprüfen. Zu diesem Zweck werden in der weiteren Analyse zwei Methoden eingesetzt, nämlich das Kaiser-Kriterium und der Bartlett-Test.

Nach dem Kaiser-Meyer-Olkin-Kriterium (Kaiser-Meyer-Olkin-Criteria, abgekürzt als KMO) kann folgendes Maß als Kennzahl für die Eignung zur Faktorenanalyse angewendet werden:

$$KMO = \frac{\sum \sum r_{ij}^2}{\sum \sum r_{ij}^2 + \sum \sum a_{ij}^2} \text{ für } i \neq j \quad (13)$$

Dabei ist  $r_{ij}^2$  die Korrelation zwischen i und j,  $a_{ij}^2$  ist die partielle Korrelation zwischen i und j.<sup>8</sup>

Die Wertbereiche von KMO können wie folgt bewertet werden:

Wertbereich	Beurteilung
[ 0,9 ; 1,0 ]	fabelhaft
[ 0,8 ; 0,9 )	recht gut
[ 0,7 ; 0,8 )	mittelprächtigt
[ 0,6 ; 0,7 )	mäßig
[ 0,5 ; 0,6 )	schlecht
[ 0 ; 0,5 )	inakzeptabel

**Tabelle 2: Bewertung der Ergebnisse von dem KMO-Maß<sup>9</sup>**

Erst nach der Aufstellung der Korrelationsmatrizen und Überprüfung der Eignung zur Faktorenanalyse kann mit der Ermittlung der Faktoren begonnen werden. Zur Extrahierung der Faktoren, stehen verschiedene Verfahren zur Verfügung. Als Beispiele zählen die Hauptkomponentenmethode, die Hauptachsenmethode, die Maximum-Likelihood-Methode und die Ungewichtet-Kleinst-Quadrat-Methode. In SPSS wird per Voreinstellung die Hauptkomponentenmethode genutzt.

Als nächster Schritt ist die Anzahl der Faktoren zu bestimmen. Hierbei können wiederum verschiedene Methoden eingesetzt werden. In der vorliegenden Arbeit werden das Kaiser-Kriterium und den Screeplot verwendet.

Nach dem Kaiser-Kriterium soll die Anzahl der zu extrahierenden Faktoren gleich

<sup>8</sup> Vgl. Brosius, F. (2013): SPSS 21, S.797

<sup>9</sup> Vgl. Brosius, F. (2013): SPSS 21, S.798

der Anzahl der Faktoren mit einem Eigenwert größer eins sein. Der Eigenwert eines Faktors gibt an, welcher Teil der Gesamtstreuung aller beobachteten Variablen durch einen bestimmten Faktor erklärt wird. Bei einem Screeplot werden die Eigenwerte der Faktoren ihrer Größe nach in einem Koordinationssystem angeordnet. Nach der Faustregel soll die Anzahl an Faktoren ausgewählt werden, bei der die Kurve einen Knick bildet.

Bei einer erfolgreichen Faktorenanalyse sollen wenige, möglichst gut voneinander unterscheidbare Faktoren ermittelt werden, die inhaltlich sinnvoll interpretierbar sind. Deshalb sind die Faktoren in der Regel nach der Bestimmung der Faktorenanzahl zu rotieren, um die inhaltliche Interpretation der Faktoren zu erleichtern. Bei der Rotation wird die Ladungsmatrix transformiert, sodass die Faktorladung der Variablen, die untereinander eine homogene Bedeutung haben, möglichst hoch ist, während die Ladung anderer Variablen auf diesen Faktoren gering ausfällt.

### 3.4.2 Reliabilitätsanalyse

Im Rahmen einer Reliabilitätsanalyse wird die Zuverlässigkeit der Hintergrundfaktoren gemessen. Eine Maßzahl der Reliabilität ist das Cronbach's Alpha.

Das Cronbach's Alpha ist ein Maß der internen Konsistenz, deren Wert zwischen -1 und +1 liegt. Je höher das Cronbach's Alpha liegt, desto höher ist die Validität der Gesamtskala. Die Faustregel dabei lautet, dass der Wert von Cronbach's Alpha größer gleich 0,7 sein soll.<sup>10</sup>

$$\alpha = \frac{k}{k-1} \left( 1 - \frac{\sum_{i=1}^k S_i^2}{S_t^2} \right) \quad (14)$$

---

<sup>10</sup> Vgl. Brosius, F. (2013): SPSS 21, S.824-827

## 4. Kapitel: Auswertung der Ergebnisse

### 4.1 Nichtparametrische Tests

#### 4.1.1 Vergleich der Stichproben aus drei Erhebungsjahren

Für die nominale Variable Familienstand wurde der Chi-Quadrat-Test durchgeführt. Die Teststatistik in Tabelle 3 zeigt eine Signifikanz von null bei den beiden Variablen. Es deutet darauf hin, dass die Verteilungen der Stichproben aus den drei Erhebungsjahren nicht identisch sind.

Chi-Quadrat-Test	Familienstand
Chi-Wert	93,121
df	10
Asymptotische Signifikanz (2-seitig)	,000

Gruppenvariable: ERHEBUNGSJAHR

**Tabelle 3: Vergleich der Familienstand nach Erhebungsjahr**

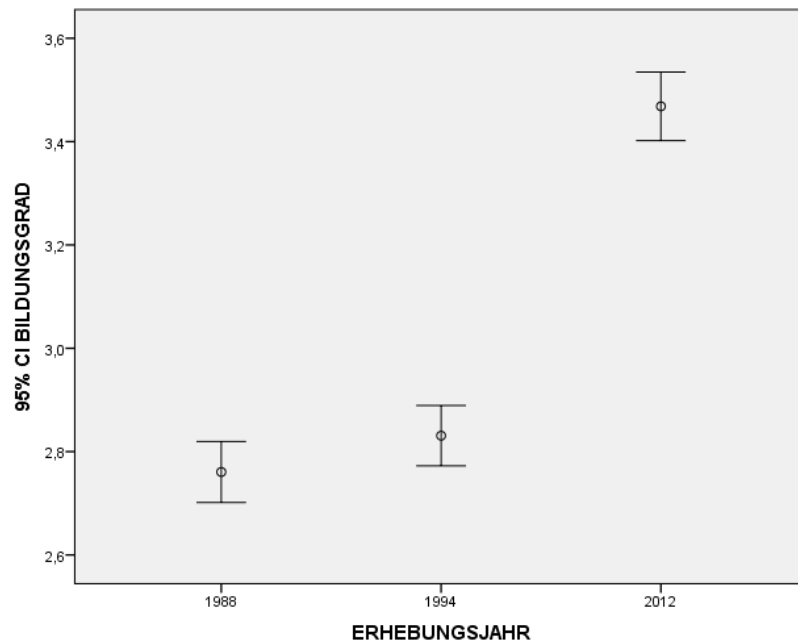
Kruskal-Wallis-Test	Bildungsgrad	Status der Berufstätigkeit	monatliches Nettoeinkommen
Chi-Wert	355,676	58,795	127,978
Df	2	2	2
Asymptotische Signifikanz (2-seitig)	,000	,000	,000

Gruppenvariable: ERHEBUNGSJAHR

**Tabelle 4: Bildung, Status der Berufstätigkeit und Einkommen nach Erhebungsjahr**

Für die Variable Bildungsgrad, Status der Erwerbstätigkeit und Einkommen wurde der Kruskal-Wallis-Test durchgeführt, die Testergebnisse sind in Tabelle 4 zu finden.

Ein Unterschied im Bildungsniveau der befragten Frauen zwischen den drei Erhebungsjahren wurde mit einer Signifikanz von null statistisch bestätigt. Das Fehlerbalkendiagramm in Abbildung 4 zeigt, dass sich die Verteilung im Jahr 2012 von der Verteilung im Jahr 1988 und 1994 signifikant unterscheidet.



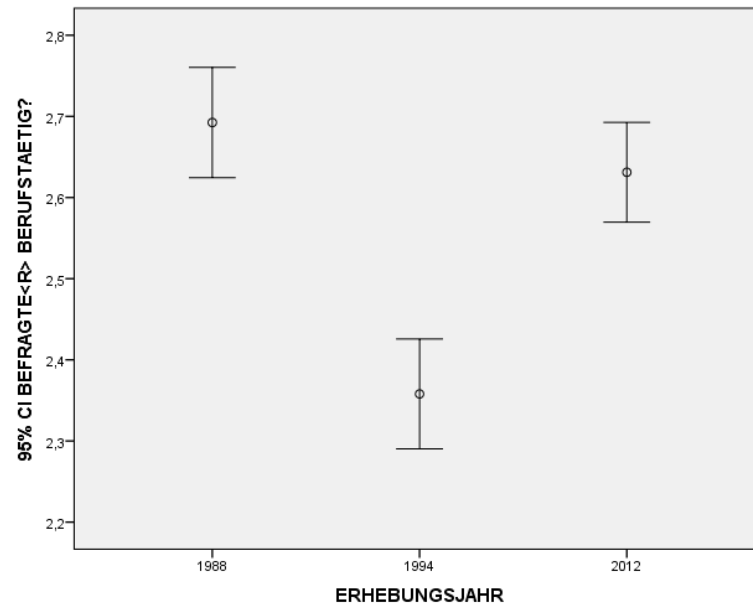
**Abbildung 4: Fehlerbalkendiagramm: Bildungsgrad nach Erhebungsjahr**

Bei der Variable Status der Erwerbstätigkeit ergibt die Teststatistik auch einen signifikanten Unterschied. In dem Fehlerbalkendiagramm in Abbildung 5 ist zu sehen, dass sich die Konfidenzintervalle der Mittelwerte der Jahre 1988 und 2012 überlappen. Das Balkendiagramm in Abbildung 2 zeigt jedoch, dass dieses Ergebnis möglicherweise deshalb zustande kommt, weil die Daten in diesen zwei Jahren bimodal verteilt sind. Obwohl die Mittelwerte und die dazugehörigen Konfidenzintervalle nahe beieinanderliegen, sind die Verteilungsformen der Daten in den beiden Jahren verschieden.

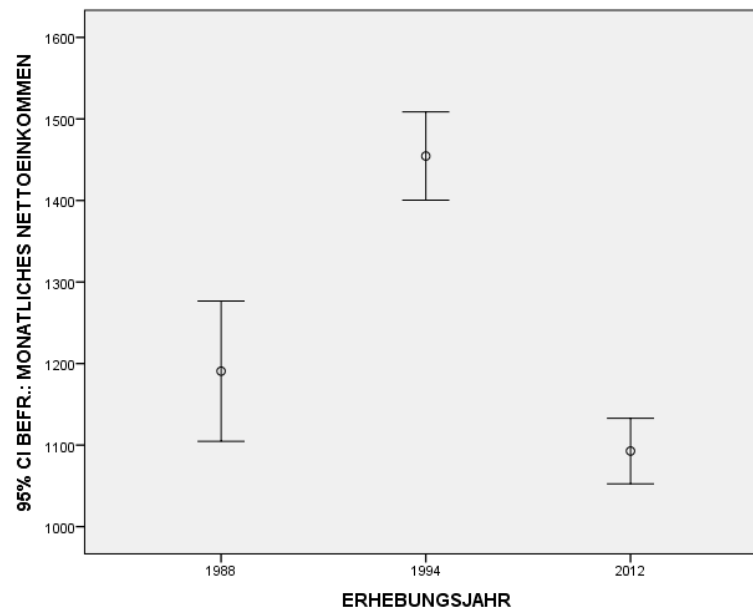
Ein Kruskal-Wallis-Test wurde ebenfalls bei der Variable Einkommen durchgeführt, weil die Voraussetzungen für ANOVA schwer verletzt sind. Mit einer Signifikanz von null wurde ein Unterschied in den Einkommensverteilungen statistisch bestätigt.



Das Fehlerbalkendiagramm in Abbildung 6 zeigt, dass sich die Einkommensverteilung im Jahr 1994 von der Verteilung aus den anderen zwei Jahren unterscheidet.



**Abbildung 5: Fehlerbalkendiagramm: Status der Berufstätigkeit nach Erhebungsjahr**



**Abbildung 6: Fehlerbalkendiagramm: Einkommen nach Erhebungsjahr**

Für die Variablen über die subjektiven Einstellungen zum Themenbereich Familie und Geschlechterrolle wurde wiederum auf den Kruskal-Wallis-Test zurückgegriffen. Die Teststatistik in Tabelle 5 legt dar, dass ein signifikanter Unterschied zwischen den Stichprobeverteilungen existiert, und zwar bei allen untersuchten Fragen. In Abbildung 2-1 bis 2-4 im Anhang ist zu sehen, dass sich die Verteilungen der Stichproben aus den drei Erhebungsjahren voneinander unterscheiden.

Kruskal-Wallis-Test	BERUFSTÄTTIGE FRAU: FAMILIENLEBEN LEIDET	FRAUEN ZIEHEN HEIM UND KIN- DER BERUF VOR	MANN GELD VERDIENEN, FRAU IM HAUS- HALT	BEI KINDER- WUNSCH SOLLTE MAN HEIRATEN
Chi-Wert	299,595	237,333	381,446	520,696
Df	2	2	2	2
Asymptotische Signifikanz (2- seitig)	,000	,000	,000	,000

Gruppenvariable: ERHEBUNGSJAHR

**Tabelle 5: Einstellungen nach Erhebungsjahr**

#### 4.1.2 Vergleich der Stichproben aus beiden Erhebungsgebieten

In dem vorliegenden Abschnitt wurden die Daten nach dem Erhebungsgebiet getrennt betrachtet und miteinander verglichen. Hierbei wurden die Daten aus den Erhebungsjahren 1994 und 2012 verwendet.

Wie in Tabelle 6 bis Tabelle 8 angezeigt ist, unterscheidet sich die Verteilung aller untersuchten Variablen in beiden Erhebungsgebieten bei den zwei Erhebungszeitpunkten signifikant voneinander. Die einzige Ausnahme bildet der Familienstand im Jahr 2012. Laut dem Ergebnis des Mann-Whitney-Tests liegt kein statistischer Unterschied in der Verteilung des Familienstands zwischen den beiden Regionen im Jahr 2012 vor.

Chi-Quadrat-Test	Familienstand	
	1994	2012
Wert	29,506	8,932
df	4	5
Asymptotische Signifikanz (2-seitig)	,000	,112

Gruppenvariable: ERHEBUNGSGEBIET

**Tabelle 6: Familienstand nach Erhebungsgebiet**

Ein Vergleich der Abbildung 2-5 mit der Abbildung 2-6 im Anhang zeigt, dass der Anteil der berufstätigen Frauen in den beiden Erhebungsgebieten zu beiden Zeitpunkten immer bei ca. 60 % liegt. Jedoch hat sich die Arbeitsform der berufstätigen Frauen hinsichtlich der Arbeitsdauer deutlich verändert. Der Anteil der berufstätigen Frauen, die eine Kurzzeitarbeit ausüben, ist in diesem Zeitraum deutlich angestiegen. Diese Tendenz ist in beiden Erhebungsgebieten zu beobachten, in den alten Bundesländern ist sie aber stärker.

Mann-Whitney-Test	Bildungsgrad		Status der Berufstätigkeit	
	1994	2012	1994	2012
Mann-Whitney-U	307925	301,333	304885	317534
Asymptotische Signifikanz (2-seitig)	,000	,002	,029	,028

Gruppenvariable: ERHEBUNGSGEBIET

**Tabelle 7: Bildung und Status der Berufstätigkeit nach Erhebungsgebiet**

In Abbildung 2-7 und 2-8 im Anhang wurde die Entwicklung der Bildungssituation der befragten Frauen veranschaulicht. Die Bildungsexpansion, die im Abschnitt 2.3 festgestellt wurde, findet in beiden Erhebungsgebieten statt. In den alten Bundesländern liegt der Anteil der gering gebildeten Frauen zu beiden Zeitpunkten höher. Gleichzeitig ist der Anteil der hoch qualifizierten Frauen in den alten Bundesländern immer höher als in den neuen Ländern.

Kolmogorov-Smirnov-Test	Monatliches Nettoeinkommen	
	1994	2012
Kolmogorov-Smirnov-Z	3,123	1,638
Asymptotische Signifikanz (2-seitig)	,000	,009

Gruppenvariable: ERHEBUNGSGEBIET

**Tabelle 8: Einkommen nach Erhebungsgebiet**

Die Boxplots in Abbildung 2-9 und 2-10 im Anhang geben die Verteilung des monatlichen Nettoeinkommens der befragten Frauen an. Obwohl sich die Einkommensverteilung im Jahr 2012 immer noch signifikant unterscheidet, hat sich die Einkommenssituation in beiden Regionen nach der Wiedervereinigung angenähert. Insbesondere ist zu sehen, dass in der zweiten Hälfte die Einkommensverteilung in den neuen Bundesländern deutlich nach oben gegangen ist. Die Anzahl der Ausreißer ist in den neuen Ländern ebenfalls gestiegen.

In Abbildung 2-11 bis 2-18 im Anhang ist zu sehen, dass sich bei den Fragen zum Thema Familie und Geschlechterrolle die Situationen der beiden Erhebungsgebiete nach der Wiedervereinigung deutlich ähneln. Von Anfang der 90er Jahre an sind modernere Einstellungen zur Familien- und Geschlechterrolle in den neuen Bundesländern viel stärker vertreten. Im Vergleich dazu fallen die Meinungen in den alten Bundesländern am Anfang der 90er Jahre gemischt aus. Jedoch ist der Anteil der Frauen in den alten Bundesländern, die einer modernen Einstellung zustimmen, in den letzten 20 Jahren deutlich gestiegen.

## 4.2 Clusteranalyse

In diesem Abschnitt soll eine Two-Step-Clusteranalyse getrennt nach dem Erhebungsjahr durchgeführt werden. Die Analyse wurde anhand folgender Merkmale untersucht: Alter, Familienstand, Bildungsniveau, Erwerbsstatus, Einkommen sowie

der Meinung zur Vereinbarkeit von Familienleben und Erwerbstätigkeit. Der Output der Clusteranalyse ist im Anhang zu finden.

Für die Daten aus Jahr 2012 wurden die befragten Frauen in zwei Cluster geteilt, der entsprechende Wert der Silhouette beträgt ca. 0,4, d. h., die Struktur der Daten ist eher schwach.

Das erste Cluster enthält ca. 30 % der gültigen Fälle. Im Durchschnitt haben die Frauen in diesem Cluster ein Alter von 68 Jahren und ein monatliches Nettoeinkommen von 804 Euro. Über 2/3 von ihnen verfügen mindestens über einen Hauptschulabschluss, ihre Meinungen zu der Vereinbarkeit von Familienleben und Erwerbstätigkeit sind gemischt.

Dem zweiten Cluster wurden ca. 70 % der gültigen Fälle zugeordnet. Im Vergleich zu dem ersten Cluster sind Frauen in dem zweiten Cluster im Durchschnitt jünger. Sie verdienen im Mittel auch ein höheres monatliches Nettoeinkommen. Mehr als 80 % von ihnen besitzen mindestens die Mittlere Reife. Über die Hälfte von ihnen hat die Aussage abgelehnt, dass das Familienleben unter der Berufstätigkeit leidet.

Für die Daten aus den Jahren 1992 und 1988 wurde auch jeweils eine Zwei-Cluster-Lösung herausgefunden, jedoch liegt der Wert der Silhouette in beiden nahe bei 0,25. Wenn sich der Wert der Silhouette in diesem Bereich befindet, kann nicht davon ausgegangen werden, dass eine Struktur in den Daten vorhanden ist.

### 4.3 Korrelationsanalyse

In diesem Abschnitt soll der Zusammenhang zwischen den betrachteten Variablen untersucht werden. Das Ergebnis wurde in Tabelle 2-9 im Anhang aufgezeigt. Hierbei wird sich in erster Linie für die Frage interessiert, wie die sozioökonomischen

Indikatoren mit den subjektiven Einstellungen zusammenhängen. Gleichzeitig ergibt sich die Frage, ob die Einstellungen der befragten Frauen zu dem untersuchten Thema von der Zeit und dem Wohngebiet abhängig sind. Darüber hinaus werden die Beziehungen zwischen den einzelnen Items über die Einstellungen zur Familie und Geschlechterrolle untersucht.

	BERUFSTÄ- TIGE FRAU: FAMILIENLE- BEN LEIDET	FRAUEN ZIE- HEN HEIM UND KINDER BERUF VOR	MANN GELD VERDIENEN, FRAU IM HAUSHALT	BEI KINDER- WUNSCH SOLLTE MAN HEIRATEN
ERHEBUNGS- JAHR	,255**	,242**	,304**	,355**
ERHEBUNGS- GEBIET	,289**	,196**	,256**	,215**
ALTER: BEFRAGTE<R>	-,197**	-,268**	-,329**	-,310**
BILDUNGSGRAD	,279**	,378**	,438**	,253**
BEFRAGTE<R> BERUFSTÄTIG?	-,023	-,055**	-,066**	-,050**
BEFR.: MONATLI- CHES NETTOEIN- KOMMEN	,080**	,147**	,149**	,042

\*\* . Die Korrelation ist auf dem 0,01 Niveau signifikant (zweiseitig).

**Tabelle 9: Auszug der Korrelationstabelle**

In Tabelle 9 ist ein Teil der Korrelationstabelle zu finden, in dem der Rangkorrelationskoeffizient zwischen den sozioökonomischen Indikatoren und den einzelnen Items über Einstellungen zur Familie und Geschlechterrolle aufgezeigt wurde. Insgesamt gesehen besteht ein signifikanter Zusammenhang zwischen fast allen untersuchten Variablen. Im Vergleich zu anderen sozioökonomischen Indikatoren hängen die Einstellungen der Befragten am stärksten mit dem Bildungsniveau zusammen. Das Alter, also die Generationszugehörigkeit der Befragten, wirkt negativ auf den ihrer Einstellungen der Befragten zum Thema Geschlechterrolle und Familie aus.

Die Zeit und das Wohngebiet üben einen größeren Einfluss auf die Einstellungen zu dem untersuchten Thema im Vergleich zu der individuellen Einkommen- und Arbeitssituation der Befragten aus.

In Tabelle 1 wurde ein Auszug der Korrelationstabelle aufgezeigt, in der der Rangkorrelationskoeffizient zwischen den einzelnen Items über Einstellungen zur Familie und Geschlechterrolle angeführt. Ganz klar ist, dass die einzelnen Items signifikant zusammenhängen. Sehr wahrscheinlich weist es darauf hin, dass diese Items durch einen gemeinsamen Faktor beeinflusst werden.

	FRAUEN ZIEHEN HEIM UND KIN- DER BERUF VOR	BERUFSTAETIGE FRAU: FAMILIEN- LEBEN LEIDET	MANN GELD VERDIENEN, FRAU IM HAUS- HALT	BEI KINDER- WUNSCH SOLLTE MAN HEIRATEN
FRAUEN ZIEHEN HEIM UND KINDER BERUF VOR	1	,505**	,663**	,397**
BERUFSTAETIGE FRAU: FAMILIEN- LEBEN LEIDET	,505**	1	,529**	,362**
MANN GELD VER- DIENEN, FRAU IM HAUSHALT	,663**	,529**	1	,466**
BEI KINDER- WUNSCH SOLLTE MAN HEIRATEN	,397**	,362**	,466**	1

\*\* . Die Korrelation ist auf dem 0,01 Niveau signifikant (zweiseitig).

**Tabelle 10: Auszug der Korrelationstabelle**

## 4.4 Faktorenanalyse

In dem vorherigen Abschnitt wurde anhand der deskriptiven Statistik und nichtparametrischen Tests gezeigt, dass sich das Meinungsbild zur Familien- und Geschlechterrolle in dem betrachteten Zeitraum dramatisch verändert hat. In diesem Abschnitt wird mittels einer explorativen Faktorenanalyse untersucht, welche Faktoren die subjektive Meinung der befragten Frauen zur Familien- und Geschlechterrolle vertreten. Hierbei werden die Daten aus Jahr 2012 herangezogen.

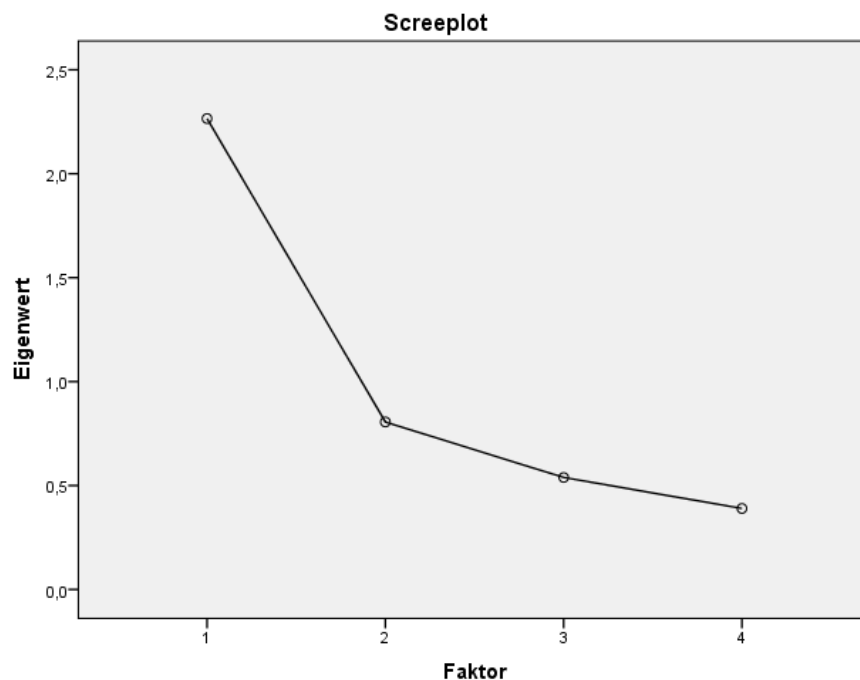
Wie in der Korrelationsmatrix in Tabelle 10 angeführt, besteht zwischen allen untersuchten Variablen eine Korrelation, die mit einer Irrtumswahrscheinlichkeit von null signifikant ist. Das KMO-Maß zeigt, dass die Daten für eine Faktorenanalyse geeignet sind.

Maß der Stichprobeneignung nach Kaiser-Meyer-Olkin.		,730
Bartlett-Test auf Sphärizität	Ungefähres Chi-Quadrat	784,701
	df	6
	Signifikanz nach Bartlett	,000

**Tabelle 11: KMO-Maß**

Als nächster Schritt soll die Anzahl der Faktoren festgelegt werden. In dem Screeplot in Abbildung 7 ist zu sehen, dass die Kurve einen Knick bildet, bei dem die Anzahl der Faktoren gleich zwei ist. Der Eigenwert ist auch größer als eins, wenn die Faktoranzahl eins beträgt. Hierbei wurde ein Ein-Faktor-Modell im SPSS ausgewählt, bei dem 56 % der Gesamtvarianz durch die extrahierten Faktoren erklärt werden. Der Anteil der erklärten Varianz ist in Tabelle 12 zu finden.





**Abbildung 7: Screeplot**

Komponente	Anfängliche Eigenwerte			Summen von quadrierten Faktorladungen für Extraktion		
	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %
1	2,265	56,635	56,635	2,265	56,635	56,635
2	,806	20,144	76,779			
3	,539	13,476	90,255			
4	,390	9,745	100,000			

Extraktionsmethode: Hauptkomponentenanalyse.

**Tabelle 12: erklärte Gesamtvarianz**

In der Komponentenmatrix in Tabelle 13 wird die Faktorladung jeder Variable aufgezeigt. Die Ladungswerte wurden mittels der Hauptkomponenten-Methode berechnet. Bei allen untersuchten Variablen sind die Ladungswerte positiv und liegen über 0,5. Mit einem Cronbach's Alpha über 0,7 gilt der Faktor als zuverlässig.

Komponentenmatrix <sup>a</sup>	
	Komponente
	1
BERUFSTAETIGE FRAU: FAMILIENLEBEN LEIDET	,770
FRAUEN ZIEHEN HEIM UND KINDER BERUF VOR	,794
MANN GELD VERDIENEN, FRAU IM HAUSHALT	,849
BEI KINDERWUNSCH SOLLTE MAN HEIRATEN	,567

Extraktionsmethode:Hauptkomponentenanalyse.

a. 1 Komponenten extrahiert

**Tabelle 13: Komponentenmatrix**

Zur Interpretation des Faktors wurden die untersuchten Items und die möglichen Antworten in Tabelle 2-10 im Anhang aufgezeigt. aufgelistet. Die genannten Items umfassen die Meinungen zu der Konsequenz der Berufstätigkeit von Frauen für ihre Familie, die Priorität zwischen Familie, Kinder, Beruf sowie die Aufgabenteilung bei der Geschlechter sowie die Notwendigkeit einer Heirat bei Kinderwunsch. Der extrahierte Faktor kann also als „Zustimmung zu dem modernen Rollenverständnis von Frauen“ interpretiert werden.

Cronbachs Alpha	Anzahl der Items
,730	4

**Tabelle 14: Cronbach's Alpha**

## 5. Kapitel: Fazit

Die wesentlichen Ergebnisse dieser Arbeit lassen sich wie folgt zusammenfassen:

- Bei allen untersuchten Variablen, die die sozioökonomischen Bedingungen der befragten Frauen und ihre Einstellungen zur Familien- und Geschlechterrolle indizieren, liegt ein statistisch signifikanter Unterschied in der Verteilung in dem betrachteten Zeitraum vor.
- Die Frauen in den beiden Bundesgebieten unterscheiden sich bei fast allen untersuchten Variablen signifikant voneinander. Im Laufe der Zeit ist der Unterschied geringer geworden.
- Im Vergleich zu den sozioökonomischen Bedingungen sind die Einstellungen der Frauen zur Familien- und Geschlechterrolle stärker von der Zeit und dem Wohngebiet geprägt.
- Anhand der Daten im Jahr 2012 lassen sich die Frauen in zwei Untergruppen differenzieren. Die erste Gruppe ist im Durchschnitt jünger, besser gebildet, verdient mehr, die Zustimmung für die Vereinbarkeit von Beruf und Familie liegt höher. Die zweite Gruppe dagegen ist im Durchschnitt älter, geringer gebildet und verdient weniger, die Zustimmung für die Vereinbarkeit von Beruf und Familie fällt niedriger aus.
- Die Einstellungen der Frauen zur Familien- und Geschlechterrolle können zu einem gemeinsamen Faktor „Zustimmung zu dem modernen Rollenverständnis für Frauen“ zusammengefasst werden.

# Literaturverzeichnis

Allerbeck K., et al.: ALLBUS 1988-Codebuch, Zentralarchiv für empirische Sozialforschung an der Universität zu Köln, Zentrum für Umfragen, Methoden und Analysen (ZUMA)

Allerbeck K., et al.: ALLBUS 1994-Codebuch, Zentralarchiv für empirische Sozialforschung an der Universität zu Köln, Zentrum für Umfragen, Methoden und Analysen (ZUMA)

Backhaus K., et al (2012): Multivariate Analysemethoden – Eine anwendungsorientierte Einführung, 13. Auflage, Berlin u.a.: Springer Verlag

Bortz, J., Lienert, G.A., Boehnke, K. (1990): Verteilungsfreie Methoden in der Biostatistik, 5. Auflage, Berlin u.a.: Springer Verlag

Bortz, J., Schuster C. (2010): Statistik Für Human- und Sozialwissenschaftler, Berlin u.a.: Springer Verlag

Brosius, F. (2013): SPSS 21, mitp

Büning, H., Trenkler, G.(1994): Nichtparametrische statistische Methoden, 2. Auflage, Berlin u.a.: Walter de Gruyter

Chiu, T., et al (1999). A Robust and Scalable Clustering Algorithm for Mixed Type Attributes in Large Database Environment. Proceedings of the seventh ACM SIGMOD international conference on knowledge discovery and data

mining, San Francisco, Kalifornien

Siegel, S.(1956): Nonparametric statistics for the behavioral sciences, New York:  
McGraw-Hill

SPSS (2001): The SPSS Two Step Cluster Component. White paper – technical  
report

TeachWiki: Die Two-Step Clusteranalyse unter SPSS

[http://mars.wiwi.hu-berlin.de/mediawiki/teachwiki/index.php/Die\\_Two-  
Step\\_Clusteranalyse\\_unter\\_SPSS](http://mars.wiwi.hu-berlin.de/mediawiki/teachwiki/index.php/Die_Two-Step_Clusteranalyse_unter_SPSS)  
(abgerufen am 29.08.2015)

Terwey M., et al.(2013): ALLBUS 2012 - Variable Report, GESIS Datenarchiv für  
Sozialwissenschaften

Träger, J.(2009): Familie im Umbruch: Quantitative und qualitative Befunde zur  
Wahl von Familienmodellen, Wiesbaden: VS Verlag für Sozialwissen-  
schaften

# Anhang: Tabellen und Abbildungen

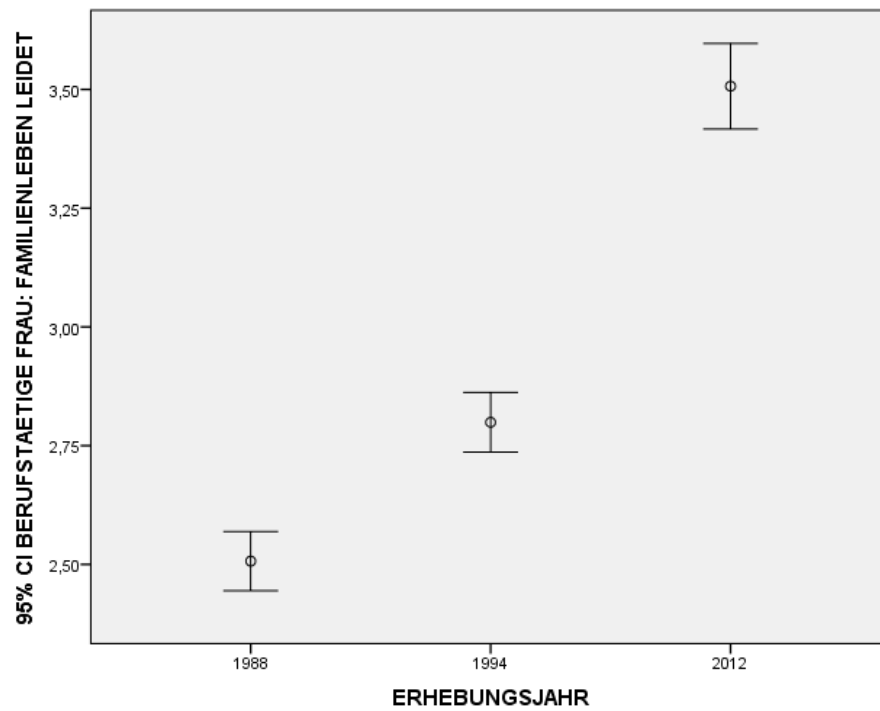
Tabelle 2-1: deskriptive Statistik der ökonomischen Indikatoren

	Alter			Familienstand			Status der Berufstätigkeit			Bildungsgrad			Einkommen		
	1988	1994	2012	1988	1994	2012	1988	1994	2012	1988	1994	2012	1988	1994	2012
<b>Gütig</b>	1696	1707	1748	1696	1709	1755	1694	1704	1754	1662	1685	1734	265	951	1237
<b>Fehlend</b>	0	2	7	0	0	0	2	5	1	34	24	21	1431	758	518
<b>Mittel- wert</b>	46.34	47.31	49.65	2.38	2.17	2.46	2.69	2.36	2.63	2.76	2.83	3.47	1190.69	1454.49	1092.71
<b>Median</b>	45	45	49	1	1	1	4	2	3	2	2	3	1000	1250	930
<b>SA</b>	18.47	17.54	17.73	1.64	1.58	1.73	1.43	1.42	1.31	1.23	1.22	1.41	711.07	850.09	721.22
<b>Mini- mum</b>	18	18	18	1	1	1	1	1	1	1	1	1	120	80	30
<b>Maxi- mum</b>	94	93	96	5	5	6	4	4	4	6	6	6	5000	6000	5000
<b>Q0,25</b>	30	32	36	1	1	1	1	1	1	2	2	2	630	850	600
<b>Q0,5</b>	45	45	49	1	1	1	4	2	3	2	2	3	1000	1250	930
<b>Q0,75</b>	62	60	63	4	3	4	4	4	4	3	3	6	1500	2000	1500

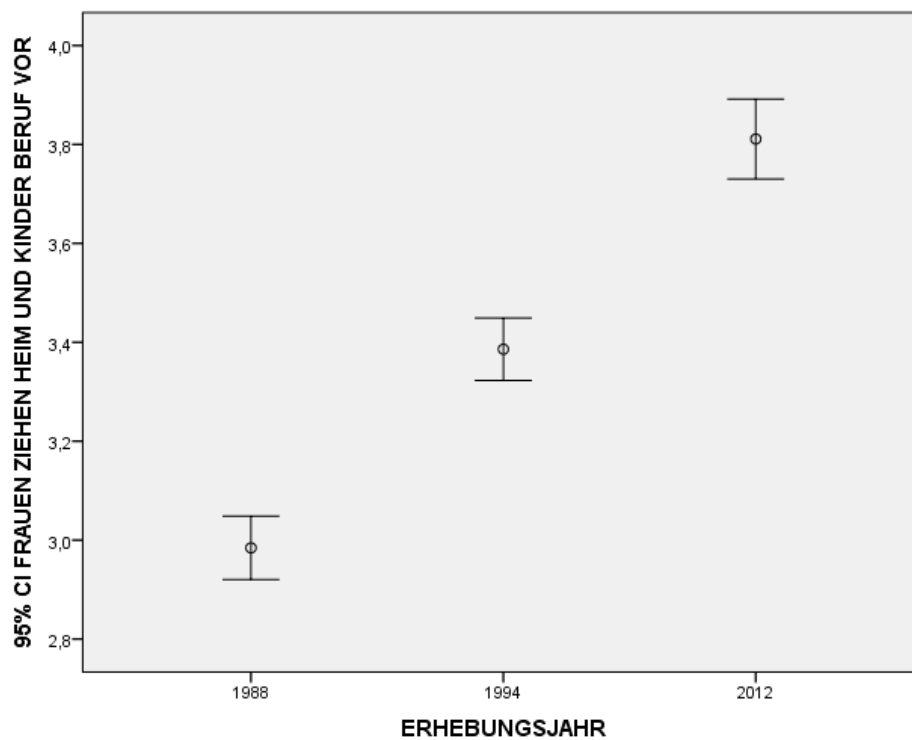
**Tabelle 2-2: deskriptive Statistik**  
**der Items über Einstellungen zur Familien- und Geschlechterrolle**

	berufstätige Frau: Familienleben leidet			Frauen ziehen Heim und Kinder Beruf vor			Mann Geld verdienen, Frauen im Haushalt			bei Kinderwunsch sollte man heiraten		
	1988	1994	2012	1988	1994	2012	1988	1994	2012	1988	1994	2012
<b>Gültig</b>	1590	1648	884	1530	1601	868	1587	1653	889	1597	1634	889
<b>Fehlend</b>	106	61	871	166	108	887	109	56	866	99	75	866
<b>Mittelwert</b>	2,51	2,80	3,51	2,98	3,39	3,81	2,92	3,53	4,01	2,01	2,63	3,32
<b>Median</b>	2	2	4	3	4	4	3	4	4	2	2	4
<b>SA</b>	1.27	1.3	1.37	1.28	1.29	1.21	1.39	1.3	1.65	1.19	1.36	1.33
<b>Q0,25</b>	2	2	2	2	2	3	2	2	4	1	2	2
<b>Q0,5</b>	2	2	4	3	4	4	3	4	4	2	2	4
<b>Q0,75</b>	4	4	5	4	4	5	4	5	5	3	4	4

**Abbildung 2-1: Fehlerbalkendiagramm**  
**Berufstätige Frau: Familienleben leidet**

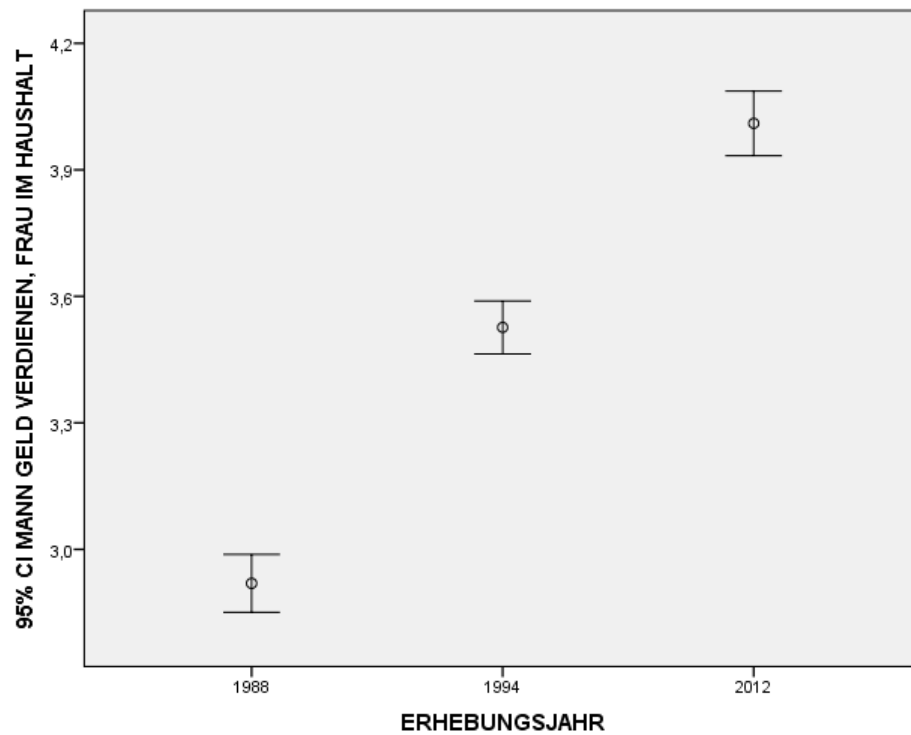


**Abbildung 2-2: Fehlerbalkendiagramm**  
**Frauen ziehen Heim und Kinder Beruf vor**

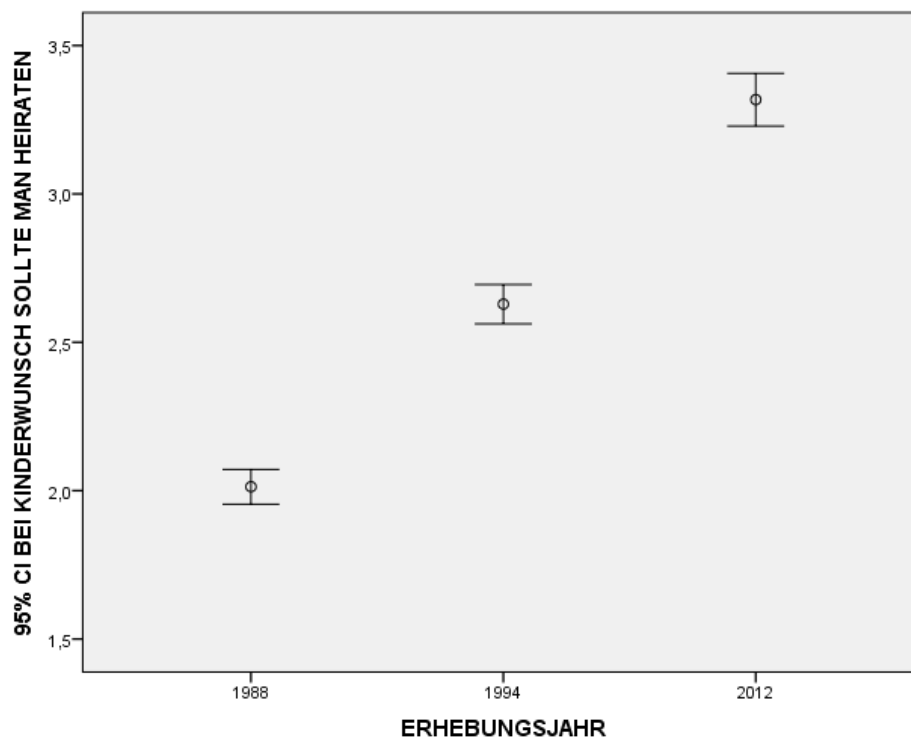




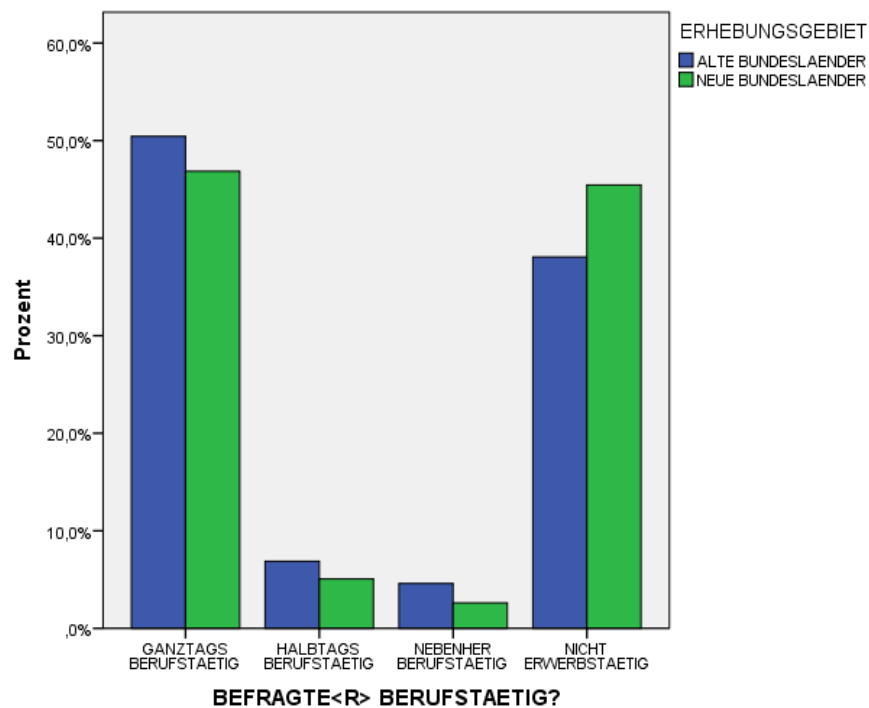
**Abbildung 2-3: Fehlerbalkendiagramm  
Mann Geld verdienen, Frau im Haushalt**



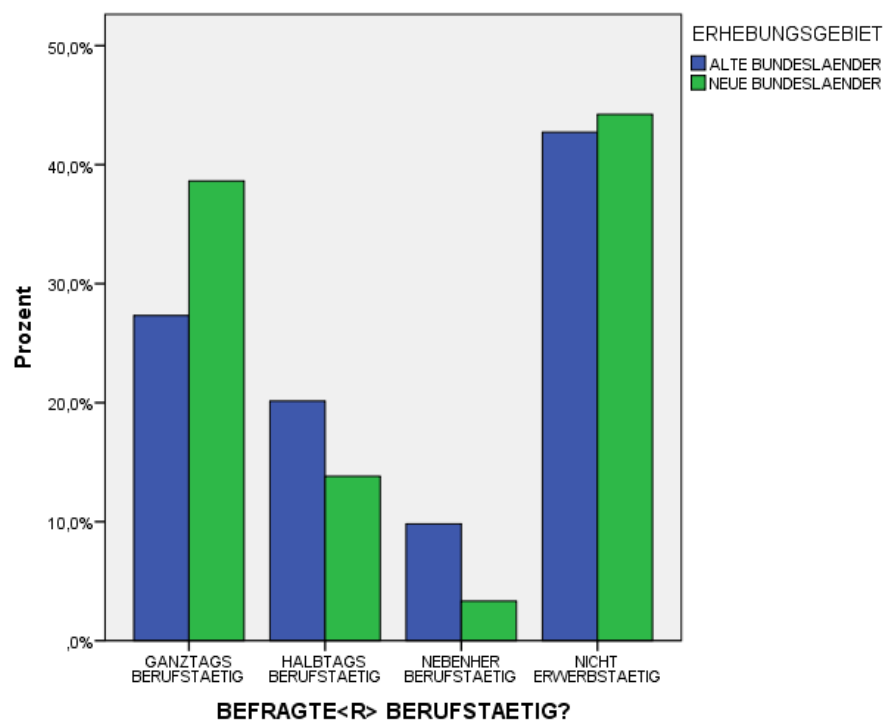
**Abbildung 2-4: Fehlerbalkendiagramm  
Bei Kinderwunsch sollte man heiraten**



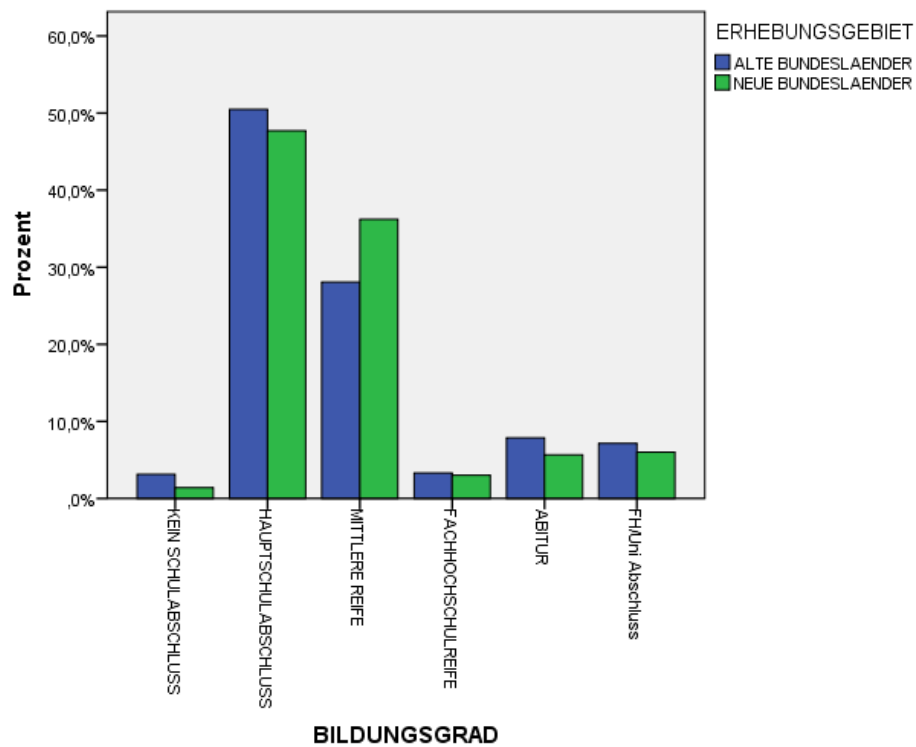
**Abbildung 2-5: Berufstätigkeit nach Erhebungsgebiet,  
Erhebungsjahr 1994**



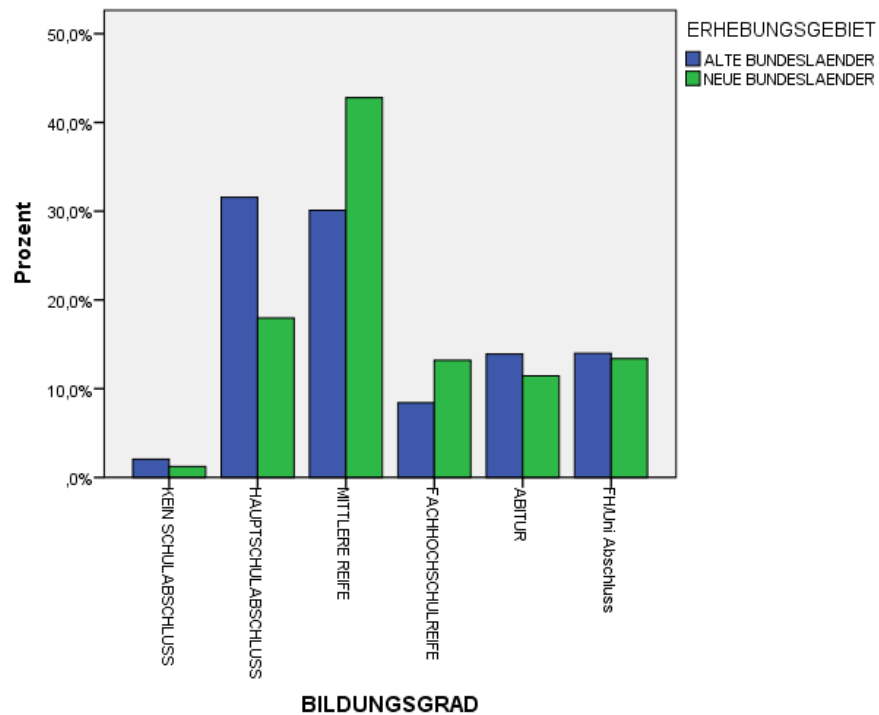
**Abbildung 2-6: Berufstätigkeit nach Erhebungsgebiet,  
Erhebungsjahr 2012**



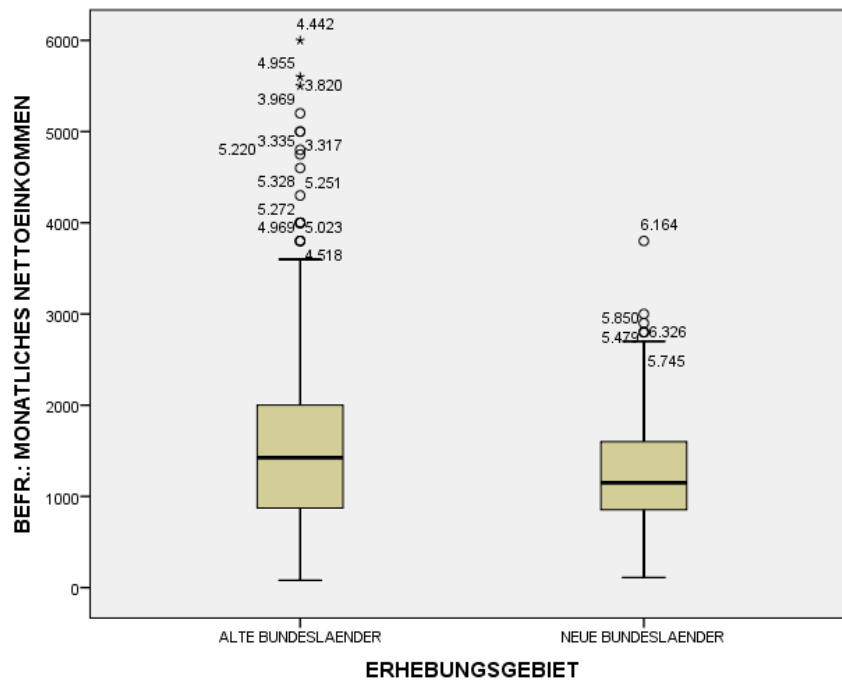
**Abbildung 2-7: Bildungsgrad nach Erhebungsgebiet,  
Erhebungsjahr 1994**



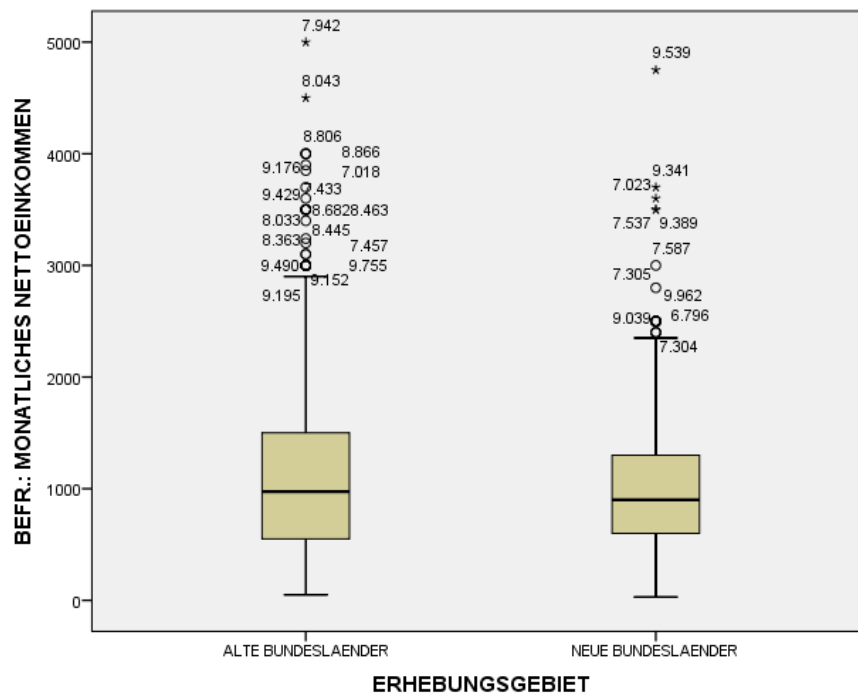
**Abbildung 2-8: Bildungsgrad nach Erhebungsgebiet,  
Erhebungsjahr 2012**



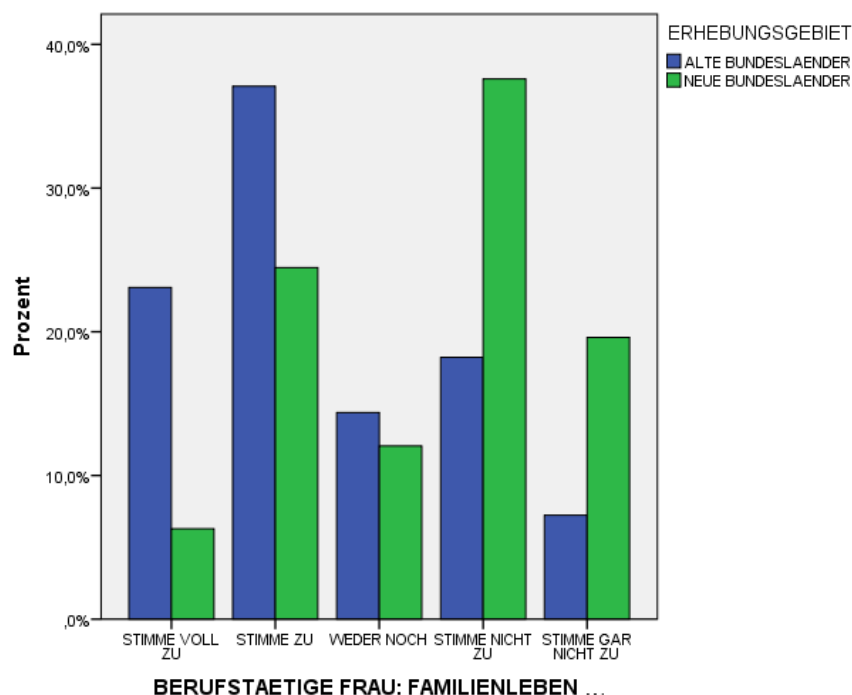
**Abbildung 2-9: Einkommen nach Erhebungsgebiet,  
Erhebungsjahr 1994**



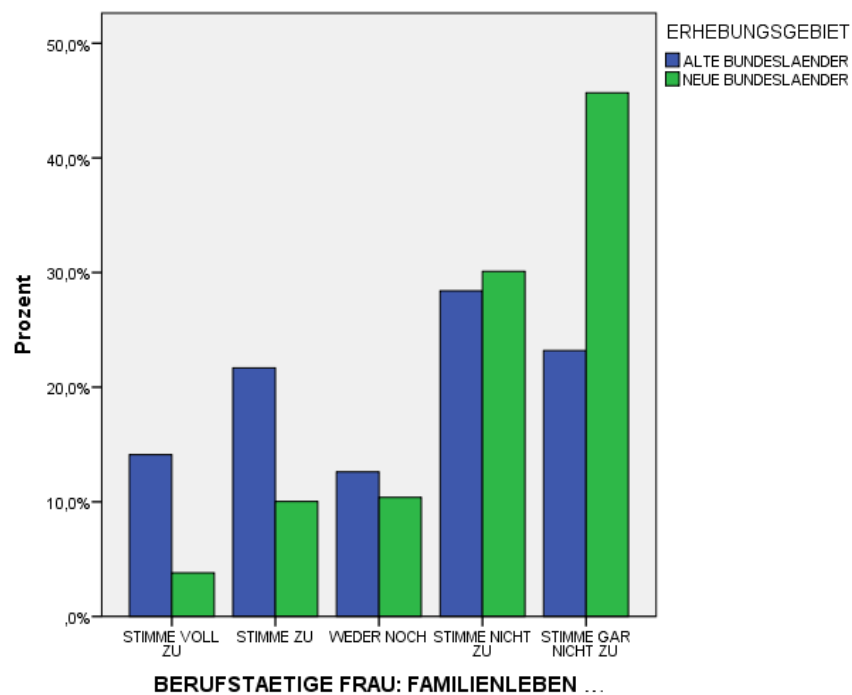
**Abbildung 2-10: Einkommen nach Erhebungsgebiet,  
Erhebungsjahr 2012**



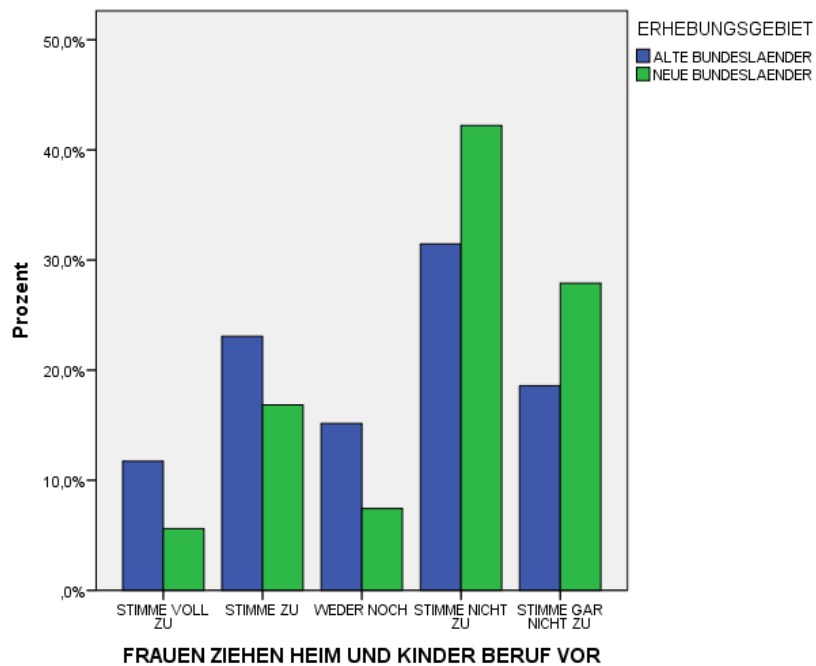
**Abbildung 2-11: Item „Berufstätige Frau: Familienleben leidet“  
Erhebungsjahr 1994, nach Erhebungsgebiet**



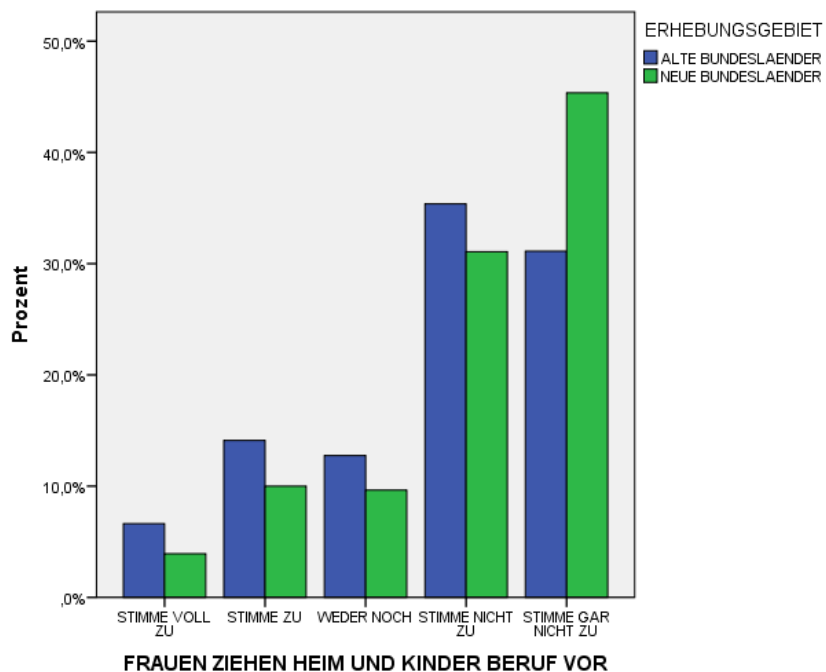
**Abbildung 2-12: Item „Berufstätige Frau: Familienleben leidet“  
Erhebungsjahr 2012, nach Erhebungsgebiet**



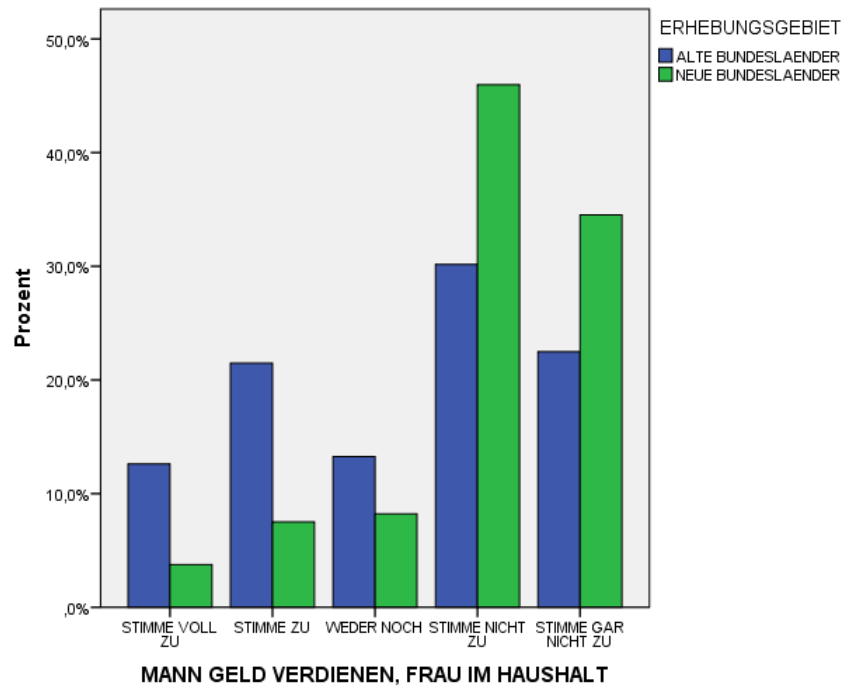
**Abbildung 2-13: Item „Frauen ziehen Heim und Kinder Beruf vor“  
Erhebungsjahr 1994, nach Erhebungsgebiet**



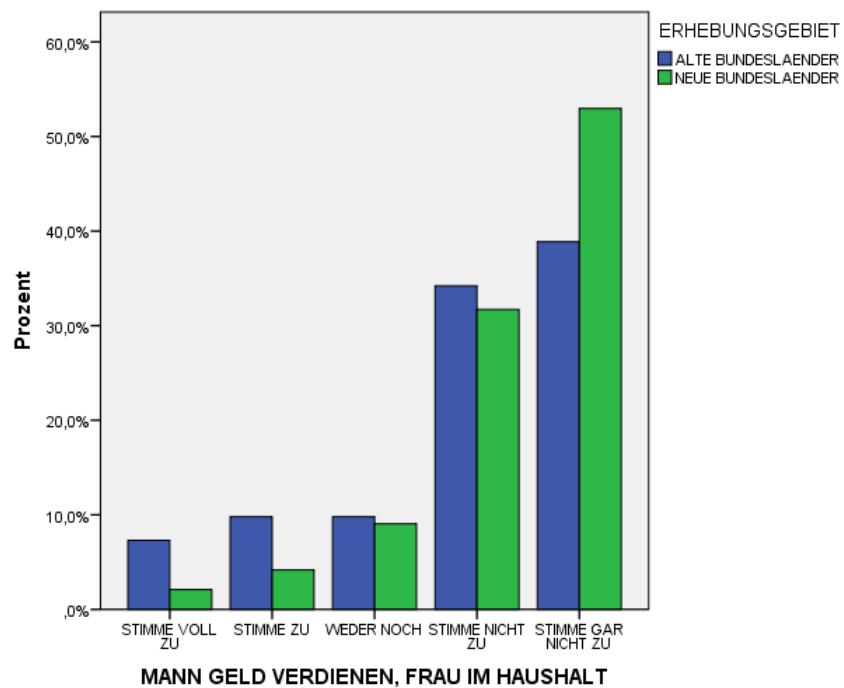
**Abbildung 2-14: Item „Frauen ziehen Heim und Kinder Beruf vor“  
Erhebungsjahr 2012, nach Erhebungsgebiet**



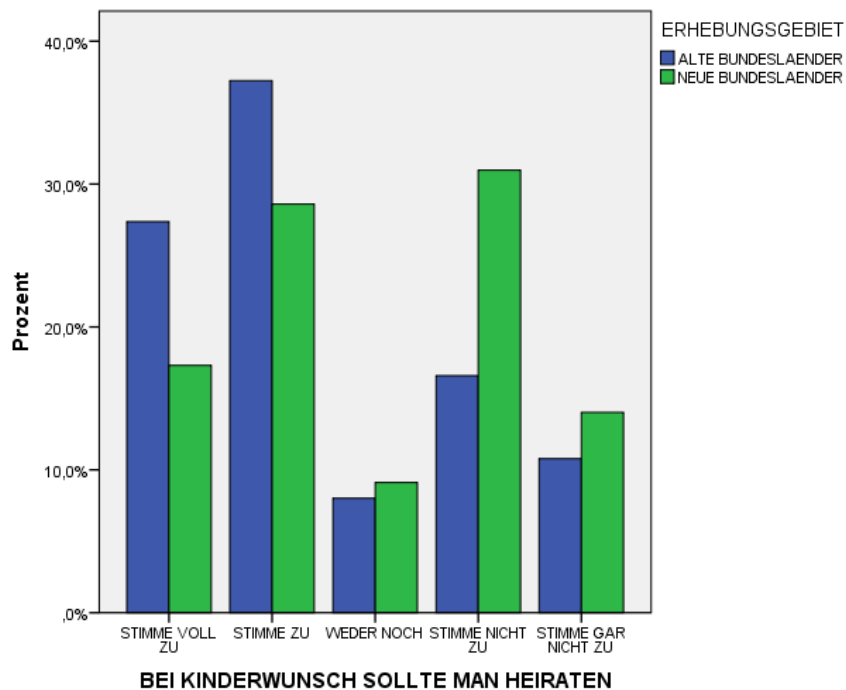
**Abbildung 2-15: Item „Man Geld verdienen, Frau im Haushalt“  
Erhebungsjahr 1994, nach Erhebungsgebiet**



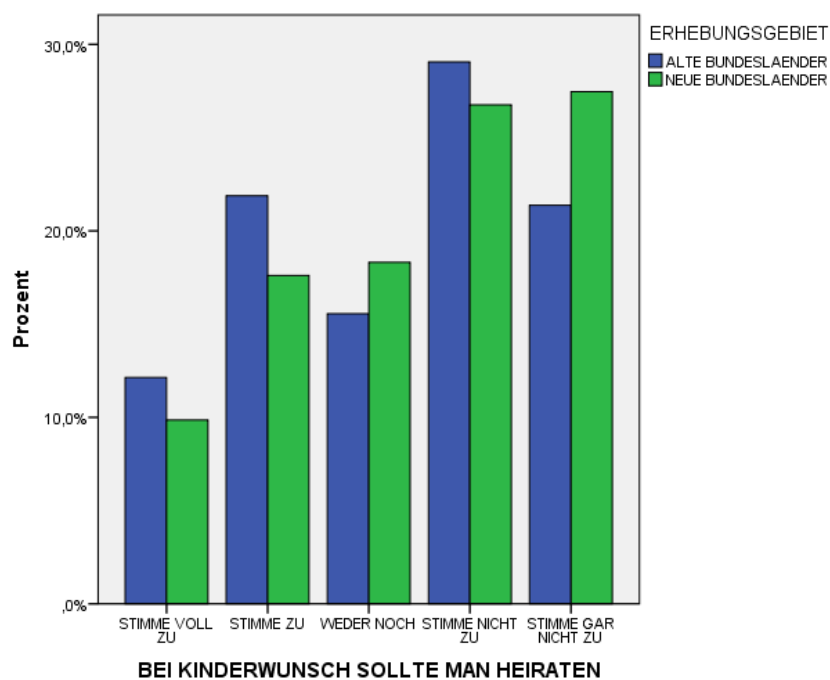
**Abbildung 2-16: Item „Man Geld verdienen, Frau im Haushalt“  
Erhebungsjahr 2012, nach Erhebungsgebiet**



**Abbildung 2-17: Item „bei Kinderwunsch sollte man heiraten“  
Erhebungsjahr 1994, nach Erhebungsgebiet**

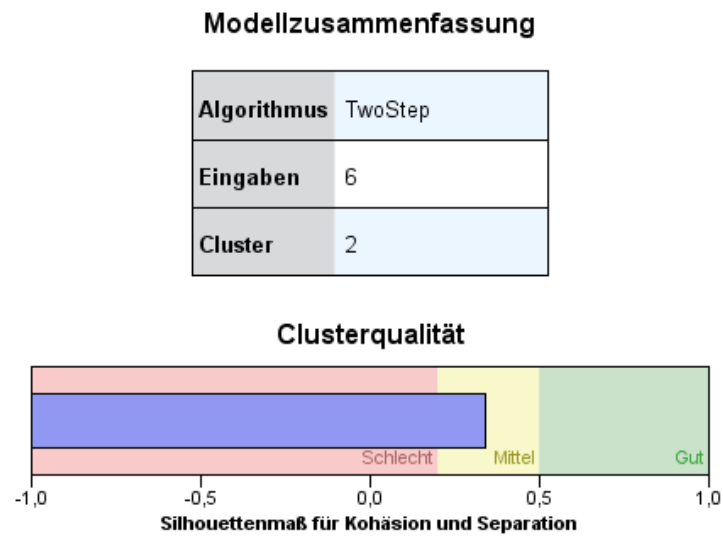


**Abbildung 2-18: Item „bei Kinderwunsch sollte man heiraten“  
Erhebungsjahr 2012, nach Erhebungsgebiet**





**Abbildung 2-19: Modellzusammenfassung und Clusterqualität**



**Tabelle 2-3: Clusterverteilung**

		N	% der Kombination	% der Gesamtsumme
Cluster	1	185	29,8%	10,5%
	2	436	70,2%	24,8%
	Kombiniert	621	100,0%	35,4%
Ausgeschlossene Fälle		1134		64,6%
Gesamtwert		1755		100,0%

**Tabelle 2-4: Zentroide**

		BEFR.: MONATLICHES NET- TOEINKOMMEN		ALTER: BEFRAGTE<R>	
		Mittelwert	Standardabweichung	Mittelwert	Standardabweichung
Cluster	1	804,00	429,393	68,89	11,506
	2	1223,94	799,649	41,45	13,032
	Kombiniert	1098,84	735,052	49,63	17,782

**Tabelle 2-5: FAMILIENSTAND, BEFRAGTE<R>**

	VERHEIRATET,ZUSAMMEN		VERHEIRATET,GETRENNT		VERWITWET	
	Häufigkeit	Prozent	Häufigkeit	Prozent	Häufigkeit	Prozent
Cluster 1	102	32,8%	0	0,0%	53	82,8%
Cluster 2	209	67,2%	7	100,0%	11	17,2%
Kombiniert	311	100,0%	7	100,0%	64	100,0%

GESCHIEDEN		LEDIG		LEBENSPARTNER	
Häufigkeit	Prozent	Häufigkeit	Prozent	Häufigkeit	Prozent
26	37,7%	4	2,4%	0	0,0%
43	62,3%	165	97,6%	1	100,0%
69	100,0%	169	100,0%	1	100,0%

**Tabelle 2-6: BILDUNGSGRAD**

	KEIN SCHULABSCHLUSS		HAUPTSCHULABSCHLUSS		MITTLERE REIFE	
	Häufigkeit	Prozent	Häufigkeit	Prozent	Häufigkeit	Prozent
Cluster 1	5	50,0%	118	65,2%	44	20,8%
Cluster 2	5	50,0%	63	34,8%	168	79,2%
Kombiniert	10	100,0%	181	100,0%	212	100,0%

FACHHOCHSCHULREIFE		ABITUR		FH/Uni Abschluss	
Häufigkeit	Prozent	Häufigkeit	Prozent	Häufigkeit	Prozent
7	11,9%	7	9,3%	4	4,8%
52	88,1%	68	90,7%	80	95,2%
59	100,0%	75	100,0%	84	100,0%

**Tabelle 2-7: BERUFSTAETIGE FRAU: FAMILIENLEBEN LEIDET**

	STIMME VOLL ZU		STIMME ZU		WEDER NOCH	
	Häufigkeit	Prozent	Häufigkeit	Prozent	Häufigkeit	Prozent
Cluster 1	31	56,4%	35	38,5%	25	29,4%
Cluster 2	24	43,6%	56	61,5%	60	70,6%
Kombiniert	55	100,0%	91	100,0%	85	100,0%

STIMME NICHT ZU		STIMME GAR NICHT ZU	
		Häufigkeit	Prozent
55	28,9%	39	19,5%
135	71,1%	161	80,5%
190	100,0%	200	100,0%

**Tabelle 2-8: automatische Clusterbildung**

Anzahl der Cluster	Bayes-Kriterium nach Schwarz (BIC)	BIC-Änderung <sup>a</sup>	Verhältnis der BIC-Änderungen <sup>b</sup>	Verhältnis der Distanzmaße <sup>c</sup>
1	7890,863			
2	7059,691	-831,172	1,000	1,784
3	6653,046	-406,645	,489	1,375
4	6394,196	-258,851	,311	1,330
5	6233,105	-161,091	,194	1,182
6	6117,704	-115,401	,139	1,148
7	6034,512	-83,192	,100	1,046
8	5960,869	-73,643	,089	1,162
9	5916,348	-44,522	,054	1,130
10	5892,516	-23,832	,029	1,127
11	5886,565	-5,950	,007	1,060
12	5888,650	2,085	-,003	1,081
13	5900,706	12,056	-,015	1,038
14	5917,318	16,612	-,020	1,041
15	5938,607	21,289	-,026	1,063

- a. Die Änderungen wurden von der vorherigen Anzahl an Clustern in der Tabelle übernommen.
- b. Die Änderungsquoten sind relativ zu der Änderung an den beiden Cluster-Lösungen.
- c. Die Quoten für die Distanzmaße beruhen auf der aktuellen Anzahl der Cluster im Vergleich zur vorherigen Anzahl der Cluster.

Tabelle 2-9: Korrelationsmatrix

	ERHE- BUNGSJAHR	ERHE- BUNGSGBIET	ALTER	BILD- UNGSGRAD	BERUFS TAETIG?	MONATLICHES NETTO-EINKOM- MEN	BERUFS- TAETIGE FRAU: FAMI- LIENLEBEN LEIDET	FRAUEN ZIE- HEN HEIM UND KINDER BERUF VOR	MANN GELD VERDIENEN, FRAU IM HAUSHALT	BEI KINDER- WUNSCH SOLLTE MAN HEIRATEN
ERHEBUNGSJAHR	1	,318**	,078**	,246**	-,009	-,175**	,255**	,242**	,304**	,355**
ERHEBUNGSGBIET	,318**	1	,075**	,103**	-,022	-,053**	,289**	,196**	,256**	,215**
ALTER	,078**	,075**	1	-,346**	,155**	-,039	-,197**	-,268**	-,329**	-,310**
BILDUNGSGRAD	,246**	,103**	-,346**	1	-,089**	,228**	,279**	,378**	,438**	,253**
BERUFSTAETIG?	-,009	-,022	,155**	-,089**	1	-,238**	-,023	-,055**	-,066**	-,050**
MONATLICHES NETTOEIN- KOMMEN	-,175**	-,053**	-,039	,228**	-,238**	1	,080**	,147**	,149**	,042
BERUFSTAETIGE FRAU: FAMILIEN LEBEN LEIDET	,255**	,289**	-,197**	,279**	-,023	,080**	1	,505**	,529**	,362**
FRAUEN ZIEHEN HEIM UND KINDER BERUF VOR	,242**	,196**	-,268**	,378**	-,055**	,147**	,505**	1	,663**	,397**
MANN GELD VERDIENEN, FRAU IM HAUSHALT	,304**	,256**	-,329**	,438**	-,066**	,149**	,529**	,663**	1	,466**
BEI KINDERWUNSCH SOLLTE MAN HEIRATEN	,355**	,215**	-,310**	,253**	-,050**	,042	,362**	,397**	,466**	1

**Tabelle 2-10: Items zum Thema Familie und Geschlechterrolle**

<b>Variablen</b>	<b>Wert</b>	<b>Ausprägung</b>	<b>Aussage</b>
BERUFSTAETIGE FRAU: FAMILIEN- LEBEN LEIDET	1	Stimme voll und ganz zu	Alles in allem: Das Familienle- ben leidet darunter, wenn die Frau voll berufstätig ist
	2	Stimme zu	
	3	Weder noch	
	4	Stimme nicht zu	
	5	Stimme überhaupt nicht zu	
FRAUEN ZIEHEN HEIM UND KIN- DER BERUF VOR	1	Stimme voll und ganz zu	Einen Beruf zu haben ist ja ganz schön, aber das, was die meisten Frauen wirklich wol- len, sind ein Heim und Kinder
	2	Stimme zu	
	3	Weder noch	
	4	Stimme nicht zu	
	5	Stimme überhaupt nicht zu	
MANN GELD VERDIENEN, FRAU IM HAUS- HALT	1	Stimme voll und ganz zu	Die Aufgabe des Ehemannes ist es Geld zu verdienen, die der Ehefrau, sich um Haushalt und Familie zu kümmern
	2	Stimme zu	
	3	Weder noch	
	4	Stimme nicht zu	
	5	Stimme überhaupt nicht zu	
BEI KINDER- WUNSCH SOLLTE MAN HEIRATEN	1	Stimme voll und ganz zu	Menschen, die Kinder wollen, sollten heiraten
	2	Stimme zu	
	3	Weder noch	
	4	Stimme nicht zu	
	5	Stimme überhaupt nicht zu	

# Erklärung zur Urheberschaft

Hiermit erkläre ich, Shujun Huang, dass ich die vorliegende Arbeit allein und nur unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe.

Die Prüfungsordnung ist mir bekannt. Ich habe in meinem Studienfach bisher keine Bachelorarbeit eingereicht bzw. diese nicht endgültig nicht bestanden.

Shujun Huang

Berlin, den 31.08.2015